

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 12:50:42 ; Search time 3056 seconds
(without alignments)
11027.830 Million cell updates/sec

Title: US-09-814-661a-1

Perfect score: 1158

Sequence: 1 aatgacacacgtgtcaaca.....cctctctgatcaatgcttt 1158

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1147	99.1	20951	8	SC9958	Z46729 S.cerevisia
2	707	61.1	2398	8	SCCMP2	X54964 Yeast CMP2
3	74	6.4	1767	6	I90044	I90044 Sequence 20
4	74	6.4	2353	6	I90040	I90040 Sequence 11
5	74	6.4	2353	8	YSCCALA2	M64840 S.cerevisia
6	74	6.4	7218	6	I66494	I66494 Sequence 14
7	64	5.5	81131	2	AC094436	AC094436 Rattus no
8	63	5.4	99679	9	CNS018SP	AL121838 Human chr
9	62.6	5.4	183506	2	AC095849	AC095849 Rattus no
10	62.6	5.4	192330	2	AC122157	AC122157 Canis fam
11	61.8	5.3	108558	2	AC107244	AC107244 Rattus no
12	61.8	5.3	189315	10	AL606508	AL606508 Mouse DNA
13	61.6	5.3	169514	9	AC064869	AC064869 Homo sapi
14	60.2	5.2	138859	2	AC129452	AC129452 Rattus no
15	60	5.2	145153	2	AC101819	AC101819 Mus muscu
16	60	5.2	192929	2	AC005505	AC005505 Plasmodiu
17	60	5.2	256172	2	AC005139	AC005139 Plasmodiu
18	59.6	5.1	197946	10	AC098722	AC098722 Mus muscu
19	59.4	5.1	189123	2	AC093998	AC093998 Rattus no
20	59.4	5.1	280342	2	AC124543	AC124543 Mus muscu
21	59.2	5.1	124704	2	AL837520	AL837520 Mus muscu
22	59.2	5.1	157420	2	AL845482	AL845482 Mus muscu
23	59.2	5.1	167758	2	AC110189	AC110189 Mus muscu
24	59.2	5.1	176733	2	AC068001	AC068001 Homo sapi
25	59	5.1	159893	8	OSJN00095	AL606652 Oryza sat
26	59	5.1	203395	2	AL805906	AL805906 Mus muscu
27	58.8	5.1	114254	2	AC126432	AC126432 Mus muscu
28	58.8	5.1	221160	2	AC102535	AC102535 Mus muscu
29	58.6	5.1	167413	9	AC099571	AC099571 Homo sapi
30	58	5.0	208161	2	AC074145	AC074145 Mus muscu
31	58	5.0	224476	2	AC118389	AC118389 Rattus no
32	57.8	5.0	39922	9	AL732637	AL732637 Human DNA
33	57.8	5.0	172116	2	AC087100	AC087100 Mus muscu
34	57.6	5.0	5455	6	AX281168	AX281168 Sequence
35	57.6	5.0	5455	6	AX356419	AX356419 Sequence
36	57.6	5.0	117538	2	AC121637	AC121637 Rattus no
37	57.6	5.0	349980	6	AX344556	AX344556 Sequence
38	57.4	5.0	10097	10	RMLTYAMG	X16379 Rat gene fo
39	57.4	5.0	178533	2	AC113233	AC113233 Canis fam
40	57.4	5.0	188772	10	AL773525	AL773525 Mouse DNA
41	57.4	5.0	349980	6	AX344550	AX344550 Sequence
42	57.2	4.9	1124	9	BC025950	BC025950 Homo sapi
43	57.2	4.9	38688	2	AC128207	AC128207 Rattus no
44	57.2	4.9	58164	2	AC102169	AC102169 Mus muscu
45	57	4.9	14529	3	AE001397	AE001397 Plasmodiu

ALIGNMENTS

RESULT 1

SC9958

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

SC9958 20951 bp DNA linear PLN 11-AUG-1997
S.cerevisiae chromosome XIII cosmid 9958.

Z46729 Z71257

Z46729.1 GI:577134

calciuretin; CNA2; CYB2; cytochrome b2; delta element; GAL80;

inosine-5'-monophosphate dehydrogenase; pif1; protein phosphatase;

transfer RNA-Tyr.

Saccharomyces cerevisiae.

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 20951)

Devlin,K. and Churcher,C.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished

2 (bases 1 to 20951)
Barrell, B. and Rajandream, M.A.
Direct Submission
Submitted (19-NOV-1994) Saccharomyces cerevisiae chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall Hinxton, Cambridge
CB10 1HQ by Bart Barrell and Marie-Adele Rajandream. E-mail:
barrell@sanger.ac.uk

COMMENT

All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50% of their
length by a larger CDS have been omitted from this analysis.
Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are also
included but some of these may be fortuitous. The length in
codons/amino-acids is given for each CDS as is the calculated codon
adaptation index (CAI).

Cosmid 9958 is overlapped at the start of this sequence by cosmid
9745

and at the end of this sequence by cosmid 9827.

Bases 7395 to 7399 and 14397 to 14398 are ambiguous.

FEATURES
source

Location/Qualifiers
1..20951

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 9958"

/complement(1..507)

/gene="PIF1"

/complement(<1..507)

/gene="PIF1"

/note="YM9958.01c, incomplete PIF1 gene, len: 169, CAI:
0.14"

/codon_start=1

/product="Pif1p"

/protein_id="CAA86714.1"

/db_xref="GI:577135"

/db_xref="SWISS-PROT:P07271"

/translation="MPKWRSTLHIIIPRRPFICSENSFLKKVSHAKLSFMSRRG
FRSNFIQALQKPSILSKEDLDLSDWDEPDICOLETEKEKKIITDIHKEDPV
DKPMRDKVNFINKDPSLWDMFPIIQQPOLISENSFDQSSQKKSRSTGKFP
LRPALKRES"

846..1976

/note="YM9958.02, unknown orf, len: 376, CAI: 0.15"

/codon_start=1

/product="unknown"

/protein_id="CAA86715.1"

/db_xref="GI:577136"

/db_xref="SWISS-PROT:P53397"

/translation="MSYKFGKLAINKSELCLANVLQAGQSPRWIDKLNQYSTTMKI
QQBKSYVILRODENEILEFVAGCGNODALKTHLMKYFRDVSLSKHLFDNWIP
SDAKALSPQIRILAQEPWETLISFICSNNNISRTMCSNLSNFGNLTIDG
VAYHSFSELTSRATEAKLRLEGFGYRAKYIETARKLYNDKAEANITSDTYLQS
ICKDAQEDVREHLSYNGVGVKADVCCLMGLHMDGIVPDVHVSRARIDYQISAN
KNHLKRYTNALPISRKKNLELDHRLMLFKKWSYAGVAGQVLFSEIGTSGS
TTGTITKRRKWDMIKETEAVITVQMKLKLVELSDHIREAKID"

complement(2194..7233)

/note="YM9958.03c, similar to unknown C. elegans gene
YOLA_CAEEL_ZK370.4 CE00396, and E.coli YCHK_SW:YCHK_ECOLI
P37053 and Shigella SW:YCHK_SHIFL_P37054; len: 1679, CAI:
0.17"

/codon_start=1

/product="unknown"

/protein_id="CAA86716.1"

/db_xref="GI:577137"

/db_xref="SWISS-PROT:Q04958"

/translation="MRSMNCITNNNTGNTKNSLGSFNSSNYSYRFTCLTDQI
ISEAQWLSLNFWSVYVFMGASRMIFRYGWLATLSLLRIPKWFKLHVQF
TLFWLILFALVIVFYVTYIMKERILSQQYKRLTPEFLPLENTKSGSSANNASTQ
SANAPPAIGSTTGASSTIDSKHSLDGNENETFLSYLDQFLSAIKIFGLKPVF
HDLTKNMKTQMDGEILLDDSTIGFAIVVEGTQLQYHEVDHSDKHGDETDSDTG
LDDQDEDEDEDDDDINDYTKSCSNLDEEDSVGYIHLKNGLNQFQNLNTVKG

misc_feature

7395..7399

/note="ambiguous sequence, agaag or ggagg"

CDS

8358..8672

/note="YM9958.04, unknown orf, len: 104, CAI: 0.32"

/codon_start=1

/product="unknown"

/protein_id="CAA86717.1"

/db_xref="GI:577138"

/db_xref="SWISS-PROT:Q04964"

/translation="MQNSQDFYFQNRCCQOQOQASTLRTVTMAEFRRVPLPPMAEVPV
LSTONSGSSASASSLSEWERDLEERLNSIDHDMNNKFGSGELKSMFNGQKVEEM
DF"

9155..10969

/gene="CNA2"

CDS

9155..10969

/gene="CNA2"

/note="YM9958.05, CNA2 or CMP2 gene, protein phosphatase,
P2B2_YEAST_P14747, len: 604, CAI: 0.16"

/codon_start=1

/product="calcineurin"

/protein_id="CAA86718.1"

/db_xref="GI:577139"

/db_xref="SWISS-PROT:P14747"

/translation="MSSDAIRNTHQINAAKIIENKTERPOSSTPIDSKASTVAAN
STATETSRDLYTQLDDGRVSTNRRIMKVPATISHPVTDEELFPNGIPRHEFLRD
HFKREGLSAAQAARIIVLATELSKEPNLISVPATVCGDIGHQYFDLLKLFVGG
DPATYSYLDGVYVDRGFSFECIILYLSLKFNDFHLLRGNHECKHLTSYFTFKN
EMLHKYLDIYKCCSEFNPLPLAALANGOYLCVHGIGSPELNSLODINLNREIP
SHGLMCDLLWADPIEYDEVDLKDQTEEDIVNSKTMVPHGKMAPSDRMPVNSVRGC
SYATYRAACHFLQETGLSLIRAEADAGRYMKNTKTLGFPSSLTFLFSAPLYDT
YNNKRAALTYNNYNNIQQNMTPHPLDFMDVFTWPSLFPVFGVETMLVALTNC
TEDELENTYIEELVGTDKLQAGSEATPQATPASPQKHAISTLODEIRKALRNK
ILAVKYSRVYSLUREETNKVQLKDHNSGVLPRGALSNGVKGLEALSTFERARKHD
LINEKLPSSDELKNNKYYEKVQKQVHEHDAKNDK"

misc_feature

9755..9772

/gene="CNA2"

/note="P500125 Serine/threonine specific protein
phosphatases signature"

complement(join(11169..12283,12692..13151))

/note="YM9958.06c, orf len: 524, CAI: 0.48, similar to
inosine-5'-monophosphate dehydrogenases"

/codon_start=1

/product="putative inosine-5'-monophosphate dehydrogenase"

/protein_id="CAA86719.1"

/db_xref="GI:577140"

/db_xref="SWISS-PROT:P50094"

/translation="MSAAPLDYKKALEHLKTYSSKGLSVQELMDSTTRGCLTYNDEL
VPLGLVPPPSAVSLQKLTKTITPTFFVSSPMDTVTEDAMLYMALLGIGTIIHN
CTPKEQASMKVKMKMFENGFINSPITPTTTTVEGVKMKKFCGFPFPTEDGKCPG

KLVGLVTRDIOFLDDSLVSEVMTKNPVTGIKILTEGNEILKQKKGKLLIIVDD
 NGNLVMSLRSDKMNQNYPLAKSATTQKLLCGAAIGTIEADKERLLRVAGLDVV
 ILDSOGNSVFLNMIKIKETFPDLEIIAGNVATREOAAANLIARAGDGLRIGMSGGS
 ICITDEVACGPGPOGTATYNYVCOFANQFVPCMDGCVQNTGHTIKALALGSSVYMG
 GMLAGTTPSGEYFTRDKRLKAYRGMGSDIDMOWRTGNKNASTSRKYSSESIVLVAQ
 GVSAGVADKGSIKKIPLYLNGLSQDIDGESLTSKENVONGEVREFFRTASQAL
 EGGVHNLHSYERKLYN"
 complement(11730..11768)
 /note="ps00487 IMP dehydrogenase/GMP reductase signature"
 complement(12284..12286)
 /note="splice acceptor sequence, cag"
 complement(12396..12402)
 /note="splice branch sequence tactaac"
 complement(12686..12691)
 /note="splice donor sequence gtatgt"
 13765..14301
 /note="YM9958.07, unknown orf, len: 178, CAI: 0.15"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA86720.1"
 /db_xref="GI:577141"
 /db_xref="SWISS-PROT:Q04969"
 /translation="MSSAKPINVYSIPELNQALDEALPSVFARLNYSYALLDAKLY
 IGFSTAVAGLSFFLDKFKERDQIVTYQKLLVGAYFVLSLFWFSRFEIKETGVYVXG
 RGTRKEEIVYKTKFKENKPLYLDELVLQKKGNSKELKAKLEVNKVFNESGYLQND
 YFKWFSQHNVLDTKKNE"
 14397..14398
 /note="ambiguous sequence, ct or cc"
 complement(14508..16283)
 /gene="CYB2"
 /gene="CYB2"
 /note="YM9958.08c, CYB2 gene, CYB2 precursor, len: 591, CAI: 0.21"
 /codon_start=1
 /product="cytochrome b2 precursor"
 /protein_id="CAA86721.1"
 /db_xref="GI:577142"
 /db_xref="SWISS-PROT:Q00175"
 /translation="MLKYKPLKISKNCFAALLRASKTRLNTIRAVGSTVPKSKSFEQ
 DSRKRTQSWALRVGAILAATSSVAYLNWHNQIDNEPKLDNKKQKISPAEVAKINKP
 99.1%; Score 1147; DB 8; Length 20951;
 Best Local Similarity 99.9%; Pred. No. 7.3e-253;
 Matches 1158; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 AATGACCAACCTGTCAACAAGAGTCTCAAGACCGGCTACTTATCCCAAGATCACGT 60
 DB |||||||
 QY 61 TCCTTCTGCCAACATCATTCGCGTCAAGCTCGCGCGGCTCTTTCTGACATTGGTAAGAA 120
 DB |||||||
 QY 7869 TCCTTCTGCCAACATCATTCGCGTCAAGCTCGCGCGGCTCTTTCTGACATTGGTAAGAA 7928
 DB |||||||
 QY 121 TACTTCCAACTAAGAGATCGTTCCTCTTTTGTAGGCCCAATGATAGGAAGAACA 180
 DB |||||||
 QY 7929 TACTTCCAACTAAGAGATCGTTCCTCTTTTGTAGGCCCAATGATAGGAAGAACA 7988
 DB |||||||
 QY 181 TAGATTATAAATACGTCAGAATATAGTACATATGTTTTATGTTAGACCTCGTACATAG 240
 DB |||||||
 QY 7989 TAGATTATAAATACGTCAGAATATAGTACATATGTTTTATGTTAGACCTCGTACATAG 8048
 DB |||||||
 QY 241 GAATAATTCACCTGTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGGCTCA 300
 DB |||||||
 QY 8049 GAATAATTCACCTGTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGGCTCA 8108
 DB |||||||
 QY 301 GCCCAACCGGCTCCACTACCGCGCGGCTCGCCATTTTGGGAAGTCATCCGTCGCCAAA 360
 DB |||||||
 QY 8109 GCCCAACCGGCTCCACTACCGCGCGGCTCGCCATTTTGGGAAGTCATCCGTCGCCAAA 8168
 DB |||||||
 QY 361 AGGAATAGCCATACATATCTGTTACTGTTTGGAAACATCGCCCGTTCGCCCGGATTCGG 420
 DB |||||||
 QY 8169 AGGAATAGCCATACATATCTGTTACTGTTTGGAAACATCGCCCGTTCGCCCGGATTCGG 8228
 DB |||||||

QY 421 CCTCAGCGGGTATAAAAAAGAGATCTTTTTTTTCTGGCTGCCCTT -CCATTTTAAAT 479
 DB |||||||
 QY 8229 CCTCAGCGGGTATAAAAAAGAGATCTTTTTTTTCTGGCTGCCCTTCCATTTTAAAT 8288
 DB |||||||
 QY 480 GTCTTATCTGCTCCTTTGTGATCTTACGGTCTCAGTAACTCTCTTCAACTGCTCAATAA 539
 DB |||||||
 QY 8289 GTCTTATCTGCTCCTTTGTGATCTTACGGTCTCAGTAACTCTCTTCAACTGCTCAATAA 8348
 DB |||||||
 QY 540 TTTCCCGGTATGCARAAATTCCTCAAGACTACTTTTACGCTCAAAATCGCTGCCAACACAA 599
 DB |||||||
 QY 8349 TTTCCCGGTATGCARAAATTCCTCAAGACTACTTTTACGCTCAAAATCGCTGCCAACACAA 8408
 DB |||||||
 QY 600 CAAGCCCTTCCACATTTGCGTACCGTGACCATGGCGGAATTTAGAGGGTGCCTTTGCCA 659
 DB |||||||
 QY 8409 CAAGCCCTTCCACATTTGCGTACCGTGACCATGGCGGAATTTAGAGGGTGCCTTTGCCA 8468
 DB |||||||
 QY 660 CTTATGGGTGAGTTCCTATGTTGCTACTCAAACTCCATGGGAGCTCCGCTTCCTGCC 719
 DB |||||||
 QY 8469 CTTATGGGTGAGTTCCTATGTTGCTACTCAAACTCCATGGGAGCTCCGCTTCCTGCC 8528
 DB |||||||
 QY 720 TCCGCTTCTTCATTAGAAATGTGGAAAGAGTTCGGAGGAGACTCAACTTATCGAT 779
 DB |||||||
 QY 8529 TCCGCTTCTTCATTAGAAATGTGGAAAGAGTTCGGAGGAGACTCAACTTATCGAT 8588
 DB |||||||
 QY 780 CATGACATGAACAACAATAATTTGGTCTGCGGAACATAAATCTATGTTCAACCAAGGT 839
 DB |||||||
 QY 8589 CATGACATGAACAACAATAATTTGGTCTGCGGAACATAAATCTATGTTCAACCAAGGT 8648
 DB |||||||
 QY 840 AAGGTGCGAGAAATGGACTTCTTAAAGTTCCTTTTCACTACTCTTTTCTCTTTTCCATT 899
 DB |||||||
 QY 8649 AAGGTGCGAGAAATGGACTTCTTAAAGTTCCTTTTCACTACTCTTTTCTCTTTTCCATT 8708
 DB |||||||
 QY 900 TCCCACTAGTCTGTTCTTTTCTTCTTTAGTACACCTTCTTTTCAAGGAGACTCTGCTCT 959
 DB |||||||
 QY 8709 TCCCACTAGTCTGTTCTTTTCTTCTTTAGTACACCTTCTTTTCAAGGAGACTCTGCTCT 8768
 DB |||||||
 QY 960 ACTATTGTTGTCATCTCGAAACATCTCTCCCGTGCATTTTCTCTTCTCTTTTATATACA 1019
 DB |||||||
 QY 8769 ACTATTGTTGTCATCTCGAAACATCTCTCCCGTGCATTTTCTCTTCTCTTTTATATACA 8828
 DB |||||||
 QY 1020 TATATATATATATATATATATATGCTCTCTCTACGATTTTGTATTTCTGTCCTTTAT 1079
 DB |||||||
 QY 8829 TATATATATATATATATATATATGCTCTCTCTACGATTTTGTATTTCTGTCCTTTAT 8888
 DB |||||||
 QY 1080 CAAAGATCTCTATATATATGTTGATACAGTATGCTAGCGGAGACATGTCCTCC 1139
 DB |||||||
 QY 8889 CAAAGATCTCTATATATACGTTTGTATACAGTATGCTAGCGGAGACATGTCCTCC 8948
 DB |||||||
 QY 1140 CTCTCTTGTATCAATGCTTT 1158
 DB |||||||
 QY 8949 CTCTCTTGTATCAATGCTTT 8967
 DB |||||||

RESULT 2

SCCMP2
 LOCUS Yeast CMP2 gene for calmodulin-binding protein 2.
 DEFINITION
 X54964
 X54964.1 GI:3559
 calmodulin binding protein; CMP2 gene.
 KEYWORDS
 Saccharomyces cerevisiae.
 SOURCE
 ORGANISM
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE
 1 (bases 1 to 2998)
 AUTHORS
 Liu,Y., Ishii,S., Tokai,M., Tsutsumi,H., Ohki,O., Akada,R.,
 Tanaka,K., Tsuchiya,E., Fukui,S. and Miyakawa,T.
 TITLE
 The Saccharomyces cerevisiae genes (CMP1 and CMP2) encoding
 calmodulin-binding proteins homologous to the catalytic subunit of
 mammalian protein phosphatase 2B
 JOURNAL
 Mol. Gen. Genet. 227 (1), 52-59 (1991)
 MEDLINE
 91260679

PUBMED 1646387
 REFERENCE 2 (bases 1 to 2998)
 AUTHORS Miyakawa, T.
 TITLE Direct Submission
 JOURNAL Submitted (30-OCT-1990) Miyakawa T., Department of Fermentation Technology, Faculty of Engineering, Hiroshima University, Saijo, Higashi-Hiroshima 724, Japan
 COMMENT See X54963 for related sequence.
 FEATURES
 source
 1..2998
 /organism="Saccharomyces cerevisiae"
 /strain="A364A"
 /db_xref="taxon:4932"
 /clone_lib="lambda gt11"
 539..545
 /note="consensus alpha-factor response element"
 663..669
 /note="consensus alpha-factor response element"
 807..813
 /note="consensus alpha-factor response element"
 907..2721
 /gene="CMP2"
 907..2721
 /gene="CMP2"
 /codon_start=1
 /product="calmodulin-binding protein 2"
 /protein_id="CAA38712.1"
 /db_xref="GI:3560"
 /db_xref="SWISS-PROT:P14747"
 /translation="MSSDAIRTEQINAAIKIENKTERPQSTTPIDSKASTVAAAN
 STATSRDLTYTLDGVRVSTNRINNKVPATISHTVPTDEDFQNGIPRPEFLRD
 HFKEGKLSAAQAARIVLATLEFKEPNLSVPAPITVCGDIHQYFOLLKLFVEVG
 DPATSYLFLGVDRGSEFELIVLSKLNFDHFWLLGRNHGKHLTSYFTFN
 EMHKYNDIYEKCESFNLLPLAAMNGOYLCHVGGISPELNSLODNLNFRERIP
 SHGLCDLLWADPIEYDEVLKDTEDIVNSKTMVPHHGRMAFSRDFVNSVRG
 SYATYRAACHELQETGLLSIIIRAHEADAGVRYNKTTLGLFSPSLTLFSAFNYLDT
 YNKAALIKYNNVNINQFNMTPHPLPDMFVFTWSLPEVGEKVTMLVAIILN
 TEDELNDPVTVEELVGTDKLPQAGSEATPQATPSAPKHSILDDHRRKLRNK
 ILAVAKVSMYSVEELVGTDKLPQAGSEATPQATPSAPKHSILDDHRRKLRNK
 LINEKLPPSLDELKNEKNKYKYEKVQKVHEDAKNDK"
 2407..2475
 /gene="CMP2"
 /bound_moiety="calmodulin"
 878 a 673 c 558 g 889 t
 BASE COUNT
 ORIGIN
 Query Match 61.1%; Score 707; DB 8; Length 2998;
 Best Local Similarity 99.9%; Pred. No. 6.9e-152;
 Matches 718; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 441 GATCTTTTTCCTGGCTGTCCTT-CCATTTTAAATGCTTATCTGCTCCTTTGTG 499
 Db 1 GATCTTTTTCCTGGCTGTCCTTCCCAITTTTAAATGCTTATCTGCTCCTTTGTG 60
 QY 500 ATCTTACGCTGCTACCTCTCTTCAACTGCTCAATAATTTCCGCGCTATGCAAAATTC 559
 Db 61 ATCTTACGCTGCTACCTCTCTTCAACTGCTCAATAATTTCCGCGCTATGCAAAATTC 120
 QY 560 CCAAGACTACTTTACGCTCAAAATCGCTGCCAACAAACAAGCCCTTCCACATTCG 619
 Db 121 CCAAGACTACTTTACGCTCAAAATCGCTGCCAACAAACAAGCCCTTCCACATTCG 180
 QY 620 TACCGTGACCATGGCGGAATTTAGAAGGTCGCTTTGCCACCTATGGCTGAGTTCCTAT 679
 Db 181 TACCGTGACCATGGCGGAATTTAGAAGGTCGCTTTGCCACCTATGGCTGAGTTCCTAT 240
 QY 680 GTTGCTTACTCAAACTCCATGGGAGCTCCGCTTCTGCTCCGCTTCTTCATTAGAAAT 739
 Db 241 GTTGCTTACTCAAACTCCATGGGAGCTCCGCTTCTGCTCCGCTTCTTCATTAGAAAT 300
 QY 740 GTGGGAAAGGATTTGGAGGAGACTCAACTCTATCATGATCATGACATCAACAACAA 799
 Db 301 GTGGGAAAGGATTTGGAGGAGACTCAACTCTATCATGATCATGACATCAACAACAA 360

QY 800 ATTTGGTTCTGCGGAACATAAATCTATGTTCAACGAGGTAAAGTCCGAGGAATGACTT 859
 Db 361 ATTTGGTTCTGCGGAACATAAATCTATGTTCAACGAGGTAAAGTCCGAGGAATGACTT 420
 QY 860 CTAAAGTTCTCTTTCATACATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 919
 Db 421 CTAAAGTTCTCTTTCATACATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 480
 QY 920 TCTTCTCTTAGATACCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 979
 Db 481 TCTTCTCTTAGATACCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 540
 QY 980 AACATTCTCTCCGTCGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1039
 Db 541 AACATTCTCTCCGTCGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 600
 QY 1040 TATGCTCTCTTACGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1099
 Db 601 TATGCTCTCTTACGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 660
 QY 1100 TTTGATACAGCTAGATATCGCTAGCGCCAAAGATTGTCCTCTTCTTGAATCAATGCTTT 1158
 Db 661 TTTGATACAGCTAGATATCGCTAGCGCCAAAGATTGTCCTCTTCTTGAATCAATGCTTT 719
 RESULT 3
 I90044
 LOCUS I90044 1767 bp DNA linear PAT 10-AUG-1998
 DEFINITION Sequence 20 from patent US 5723436.
 ACCESSION I90044
 VERSION I90044.1 GI:3409984
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1767)
 AUTHORS Huang, L. and Cyert, M. S.
 TITLE Calneurin interacting protein compositions and methods
 JOURNAL Patent: US 5723436-A 20 03-MAR-1998;
 FEATURES
 source
 1..1767
 /organism="unknown"
 524 a 416 c 353 g 474 t
 BASE COUNT
 ORIGIN
 Query Match 6.4%; Score 74; DB 6; Length 1767;
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1085 ATAGTCTATAATACGTTTTCATACAGCTAGATATCGCTAGCGCAACATTGTCCTCTC 1144
 Db 1 ATAGTCTATAATACGTTTTCATACAGCTAGATATCGCTAGCGCAACATTGTCCTCTC 60
 QY 1145 TTTGATCAATGCTTT 1158
 Db 61 TTTGATCAATGCTTT 74
 RESULT 4
 I90040
 LOCUS I90040 2353 bp DNA linear PAT 10-AUG-1998
 DEFINITION Sequence 11 from patent US 5723436.
 ACCESSION I90040
 VERSION I90040.1 GI:3409980
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2353)
 AUTHORS Huang, L. and Cyert, M. S.
 TITLE Calneurin interacting protein compositions and methods
 JOURNAL Patent: US 5723436-A 11 03-MAR-1998;

FEATURES source Location/Qualifiers
1..2353
/organism="unknown"
BASE COUNT 728 a 511 c 461 g 653 t
ORIGIN
|||||
Query Match 6.4%; Score 74; DB 6; Length 2353;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1085 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 1144
Db 1 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 60
|||||
Qy 1145 TTGATCAATGCTTT 1158
Db 61 TTGATCAATGCTTT 74
|||||
YSCCALA2 2353 bp DNA linear PLN 27-APR-1993
DEFINITION S.cerevisiae calcineurin A2 (CNA2) gene, complete cds.
ACCESSION M64840
VERSION M64840.1 GI:171150
KEYWORDS calcineurin A2; calmodulin-binding protein; catalytic subunit;
phosphoprotein phosphatase.
SOURCE Saccharomyces cerevisiae (strain GRF88) (library: Ycp50) DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 2353)
Cyert.M.S., Kunisawa,R., Kaim,D. and Thorner,J.
Yeast has homologs (CNA1 and CNA2 gene products) of mammalian
calcineurin, a calmodulin-regulated phosphoprotein phosphatase
Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7376-7380 (1991)
91334468
PUBMED 1651503
FEATURES source Location/Qualifiers
1..2353
/organism="Saccharomyces cerevisiae"
/strain="GRF88"
/db_xref="taxon:4932"
/cell_type="alpha"
/tissue_lib="Ycp50"
262..2076
/gene="CNA2"
262..2076
/gene="CNA2"
/codon_start=1
/product="calcineurin A2"
/protein_id="AAA34466.1"
/db_xref="GI:171151"
/translation="MSSDAIRNTQINAAIKIENKTERQSSSTTPIDSKASTVAAN
STATERSDLTOYLLDGRVSTNRRINMKVPALITVHPDDELFQNGPIRHEFLRD
HFREGKLSAAQAARIATLATELFESKPNLISVPATVTCGDIGHQYFDLLKLEFVG
DPATYSYLDGVDGDSFCEFLIYLSKLNFDHFWLLRNGHECKHLTSYTFKRN
EMLHKYLDIYEKCESFNPLAALMNGQVLCVHGGISPELNSLODINLNRFREIP
SHGLMCDLLWADPTIEEYDEVLDKLTEDDIVNSKTMVPHHGKMAPSRDMFVPNSVRG
SYATYRAACHFLOETGLSLIIRAHAQDAGRYMYKNTKLTGFPSSLITLSAPNYLDT
YNNKAALKYENNMNIRQFNTPHPYLDPMDFVTSWLPFFVGEKVTMELVALINIC
TELENDPTVPIELVETDKKLPQAGKSEATPPQATSPASPKHASILDDEHRRKALRNK
ILAVKYSRYMYSVETNKKVQFLKDHNSGVPALPRGALSNGVKGLDEALSTFERARKHD
LINEKLPPSLDELKNNKYYEKVQVQVHEHDAKNDK"
BASE COUNT 728 a 511 c 461 g 653 t
ORIGIN chromosome XIII near the SUP5 locus.
Query Match 6.4%; Score 74; DB 8; Length 2353;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1085 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 1144
|||||

Db 1 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 60
Qy 1145 TTGATCAATGCTTT 1158
Db 61 TTGATCAATGCTTT 74
|||||
RESULT 6
LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 6.4%; Score 74; DB 6; Length 7218;
Best Local Similarity 3.5%; Pred. No. 1.4e-06;
Matches 11; Conservative 206; Mismatches 101; Indels 0; Gaps 0;
Qy 841 AGGTCGAGGAATGACCTTCTAAAGTCTCTTTCATACCTCTTTCTTTCTTTTCTTTTTCATTT 900
Db 1047 AGGTCGAGGAGCTGCGATYY 1106
Qy 901 CCCACTAGTCTCTCTTTCT 960
Db 1107 YY 1166
Qy 961 CTATTGTTGTCATCTCGAAACATCTCTCCGTCGATTTCTCCCTTTTATATACAT 1020
Db 1167 YY 1226
Qy 1021 AT 1080
Db 1227 YY 1286
Qy 1081 AAAGATAGTCTATATATAGCTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCTC 1140
Db 1287 YY 1346
Qy 1141 TCTCTTGATCAATGCTTT 1158
Db 1347 YYYYYYYYYYYYYYYYYYYY 1364
|||||
RESULT 7
AC094436/c 81131 bp DNA linear HTG 10-JUL-2002
LOCUS Rattus norvegicus clone CH230-4F16, *** SEQUENCING IN PROGRESS ***
DEFINITION 48 unordered pieces.
ACCESSION AC094436
VERSION AC094436.3 GI:21716476
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 81131)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimake,K., Blankenburg,K., Bonnin,D.,

Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karisson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Hawthine, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherif, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Sytek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 81131)
 Worley, K.C.

Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 81131)
 Worley, K.C.

Direct Submission
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 9, 2002 this sequence version replaced gi:17941163.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GAPG
 Center clone name: CH230-4F16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 27478 bases at least Q40
 Consensus quality: 29457 bases at least Q30
 Consensus quality: 31122 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: this is a 'working draft' sequence. It currently
 * consists of 48 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1265:	contig of 1265 bp in length
*	1365:	gap of unknown length
*	1366:	contig of 1011 bp in length
*	2376:	gap of unknown length
*	2377:	contig of 1207 bp in length
*	3683:	gap of unknown length
*	3784:	contig of 1332 bp in length
*	5115:	gap of unknown length
*	5215:	contig of 1446 bp in length
*	6661:	gap of unknown length
*	6761:	contig of 1165 bp in length
*	7926:	gap of unknown length
*	8026:	contig of 1192 bp in length
*	9219:	gap of unknown length
*	9319:	contig of 1478 bp in length
*	10795:	gap of unknown length
*	10895:	contig of 1412 bp in length
*	12308:	gap of unknown length
*	12408:	contig of 1277 bp in length
*	13685:	gap of unknown length
*	13785:	contig of 1558 bp in length
*	15343:	gap of unknown length
*	15443:	contig of 1657 bp in length
*	17100:	gap of unknown length
*	17200:	contig of 1616 bp in length
*	18816:	gap of unknown length
*	18916:	contig of 1054 bp in length
*	19970:	gap of unknown length
*	20070:	contig of 1257 bp in length
*	21327:	gap of unknown length
*	21427:	contig of 1381 bp in length
*	22908:	gap of unknown length
*	24666:	contig of 1758 bp in length
*	24766:	gap of unknown length
*	26387:	contig of 1621 bp in length
*	26487:	gap of unknown length
*	28211:	contig of 1724 bp in length
*	28311:	gap of unknown length
*	29323:	contig of 1012 bp in length
*	29423:	gap of unknown length
*	31285:	contig of 1863 bp in length
*	31385:	gap of unknown length
*	32713:	contig of 1327 bp in length
*	32813:	gap of unknown length
*	34279:	contig of 1466 bp in length
*	34379:	gap of unknown length
*	35739:	contig of 1360 bp in length
*	35839:	gap of unknown length
*	37499:	contig of 1660 bp in length
*	37599:	gap of unknown length
*	39245:	contig of 1646 bp in length
*	39345:	gap of unknown length
*	40551:	contig of 1206 bp in length
*	40651:	gap of unknown length
*	42188:	contig of 1537 bp in length
*	42288:	gap of unknown length
*	44312:	contig of 2024 bp in length
*	44412:	gap of unknown length
*	46721:	contig of 2309 bp in length
*	46821:	gap of unknown length
*	48404:	contig of 1583 bp in length
*	48504:	gap of unknown length
*	50086:	contig of 1582 bp in length
*	50186:	gap of unknown length
*	52112:	contig of 1926 bp in length
*	52212:	gap of unknown length
*	53342:	contig of 1130 bp in length
*	53442:	gap of unknown length

RESULT 9

AC095849/c

LOCUS

DEFINITION

AC095849 183506 bp DNA linear HTG 11-JUL-2002
 Rattus norvegicus clone CH230-9P8, *** SEQUENCING IN PROGRESS ***,
 51 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC095849
 HTG: HTGS_PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 183506)
 Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
 Alsbrooks S.L., Anaratunge H.C., Are J.R., Ayelle M., Banks T.,
 Barbic J., Bencon J., Blinag K., Blankenburg K., Bonnin D.,
 Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P.,
 Buhal C., Burch C., Burkett C., Burrell K.L., Byrd N.C.,
 Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
 Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
 Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
 Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
 Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,
 Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
 Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
 Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
 Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
 Correll J.H., Guevara M., Gunaratne P., Hale S., Hamilton K.,
 Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
 Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B.,
 Homsai F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
 Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
 Jacobson E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
 Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
 Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulseghe H.,
 Lozdo R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
 Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,
 Massey E., Mawhney E., McLeod M.P., Meador M., Mei G., Metzker M.,
 Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
 Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
 Nguyen N., Nickerson E., Nwokenkwo S., Ogih M., Okwuonu G.,
 Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
 Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
 Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savary G.,
 Scherer S., Scott G., Shen H., Shoostari N., Sisson I.,
 Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H.,
 Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,
 Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
 Usmani K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q.,
 Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
 Williams G., Williamson A., Wleczyk R., Wooden S., Worley K.,
 Wu C., Wu Y., Wu F., Zhou J., Zorrilla S., Nelson D.,
 Weinstein G. and Gibbs R.

Direct Submission

Unpublished

Workley K.C.

Direct Submission

Submitted (17-SEP-2001)

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

Workley K.C.

Direct Submission

Submitted (11-JUL-2002)

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 10, 2002 this sequence version replaced gi:20975931.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GPPD

Center clone name: CH230-9P8

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 145588 bases at least Q40

Consensus quality: 151723 bases at least Q30

Consensus quality: 157385 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: this is a 'working draft' sequence. It currently
 * consists of 51 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1454: contig of 1454 bp in length
 1455: gap of unknown length
 2793: contig of 1239 bp in length
 2893: gap of unknown length
 2994: contig of 1087 bp in length
 3981: gap of unknown length
 4081: contig of 1441 bp in length
 5522: gap of unknown length
 5622: contig of 1140 bp in length
 6762: gap of unknown length
 8230: contig of 1369 bp in length
 8331: gap of unknown length
 9357: contig of 1077 bp in length
 9457: gap of unknown length
 10616: contig of 1139 bp in length
 10716: gap of unknown length
 11940: contig of 1324 bp in length
 12040: gap of unknown length
 13545: contig of 1505 bp in length
 13645: gap of unknown length
 16198: contig of 2553 bp in length
 16298: gap of unknown length
 17673: contig of 1375 bp in length
 17773: gap of unknown length
 19321: contig of 1548 bp in length
 19421: gap of unknown length
 20980: contig of 1559 bp in length
 21080: gap of unknown length
 22150: contig of 1077 bp in length
 22250: gap of unknown length
 23977: contig of 1727 bp in length
 24077: gap of unknown length
 26761: contig of 2884 bp in length
 26861: gap of unknown length
 29387: contig of 2526 bp in length
 29487: gap of unknown length
 31979: contig of 2492 bp in length
 32079: gap of unknown length
 34665: contig of 2386 bp in length
 34765: gap of unknown length
 36957: contig of 2192 bp in length
 37057: gap of unknown length
 38584: contig of 1527 bp in length
 38684: gap of unknown length
 41475: contig of 2791 bp in length
 41575: gap of unknown length
 44259: contig of 2684 bp in length
 44359: gap of unknown length
 46957: contig of 2598 bp in length
 47057: gap of unknown length
 49153: contig of 2096 bp in length

REFERENCE 1 (bases 1 to 108558)

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hqsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

*	1	1529:	contig	of 1529	bp in length
*	1630	1629:	gap	of unknown	length
*	1631	3031:	contig	of 1402	bp in length
*	3032	3131:	gap	of unknown	length
*	3132	3479:	contig	of 1348	bp in length
*	4480	4579:	gap	of unknown	length
*	4580	5936:	contig	of 1357	bp in length
*	5937	6036:	gap	of unknown	length
*	6037	7154:	contig	of 1118	bp in length
*	7155	7254:	gap	of unknown	length
*	7255	8527:	contig	of 1273	bp in length
*	8528	8627:	gap	of unknown	length
*	8628	9758:	contig	of 1131	bp in length
*	9759	9858:	gap	of unknown	length
*	9859	10957:	contig	of 1059	bp in length
*	10958	11057:	gap	of unknown	length
*	11058	12473:	contig	of 1416	bp in length
*	12474	12573:	gap	of unknown	length
*	12574	13659:	contig	of 1086	bp in length
*	13660	13759:	gap	of unknown	length
*	13760	15140:	contig	of 1381	bp in length
*	15141	15240:	gap	of unknown	length
*	15241	16365:	contig	of 1125	bp in length
*	16366	16465:	gap	of unknown	length
*	16466	17602:	contig	of 1137	bp in length
*	17603	17702:	gap	of unknown	length
*	17703	19063:	contig	of 1361	bp in length
*	19064	19163:	gap	of unknown	length
*	19164	20495:	contig	of 1332	bp in length
*	20496	20595:	gap	of unknown	length
*	20596	22031:	contig	of 1436	bp in length
*	22032	22131:	gap	of unknown	length
*	22132	23570:	contig	of 1439	bp in length
*	23571	23670:	gap	of unknown	length
*	23671	25320:	contig	of 1650	bp in length
*	25321	25420:	gap	of unknown	length
*	25421	27230:	contig	of 1810	bp in length
*	27231	27330:	gap	of unknown	length
*	27331	28764:	contig	of 1434	bp in length
*	28765	28864:	gap	of unknown	length
*	28865	30433:	contig	of 1569	bp in length
*	30434	30533:	gap	of unknown	length
*	30534	32344:	contig	of 1811	bp in length
*	32345	32444:	gap	of unknown	length
*	32445	33923:	contig	of 1479	bp in length
*	33924	34023:	gap	of unknown	length
*	34024	35454:	contig	of 1431	bp in length
*	35455	35554:	gap	of unknown	length
*	35555	36778:	contig	of 1224	bp in length
*	36779	36878:	gap	of unknown	length
*	36879	38080:	contig	of 1208	bp in length
*	38087	38186:	gap	of unknown	length
*	38187	39909:	contig	of 1723	bp in length
*	39910	40093:	gap	of unknown	length
*	40010	41821:	contig	of 1812	bp in length
*	41822	41921:	gap	of unknown	length
*	41922	43642:	contig	of 1721	bp in length
*	43643	43742:	gap	of unknown	length
*	43743	44925:	contig	of 1183	bp in length
*	44926	45025:	gap	of unknown	length
*	45026	47083:	contig	of 2058	bp in length
*	47084	47183:	gap	of unknown	length
*	47184	48888:	contig	of 1705	bp in length
*	48889	48988:	gap	of unknown	length
*	48989	50809:	contig	of 1821	bp in length

*	50810	50909:	gap of	unknown length
*	50910	52857:	contig of	1948 bp in length
*	52858	52957:	gap of	unknown length
*	52958	55047:	contig of	2090 bp in length
*	55048	55147:	gap of	unknown length
*	55148	57283:	contig of	2136 bp in length
*	57284	57383:	gap of	unknown length
*	57384	59575:	contig of	2192 bp in length
*	59576	59675:	gap of	unknown length
*	59676	61070:	contig of	1395 bp in length
*	61071	61170:	gap of	unknown length
*	61171	62744:	contig of	1574 bp in length
*	62745	62844:	gap of	unknown length
*	62845	64761:	contig of	1917 bp in length
*	64762	64861:	gap of	unknown length
*	64862	67131:	contig of	2270 bp in length
*	67132	67231:	gap of	unknown length
*	67232	69907:	contig of	2676 bp in length
*	69908	70007:	gap of	unknown length
*	70008	71336:	contig of	1329 bp in length
*	71337	71436:	gap of	unknown length
*	71437	72766:	contig of	1330 bp in length
*	72767	72866:	gap of	unknown length
*	72867	74382:	contig of	1516 bp in length
*	74383	74482:	gap of	unknown length
*	74483	76757:	contig of	2275 bp in length
*	76758	76857:	gap of	unknown length
*	76858	78901:	contig of	2044 bp in length
*	78902	79001:	gap of	unknown length
*	79002	80862:	contig of	1861 bp in length
*	80863	80962:	gap of	unknown length
*	80963	83348:	contig of	2586 bp in length
*	83549	83648:	gap of	unknown length
*	83649	85781:	contig of	2133 bp in length
*	85782	85881:	gap of	unknown length
*	85882	87715:	contig of	1834 bp in length
*	87716	87815:	gap of	unknown length
*	87816	89653:	contig of	1838 bp in length
*	89654	91815:	contig of	2062 bp in length
*	89754	91915:	gap of	unknown length
*	91816	94217:	contig of	2302 bp in length
*	91916			

RESULT 12

AL606508
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

Polymorphisms have been identified between AC064869 and AC013454.

FEATURES

Source

Location/Qualifiers

1. 169514

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone_lib="RPCI-11"

152. 464

/rpt_family="L2"

738. 1271

/rpt_family="ERV1"

1948. 2300

/rpt_family="MaLR"

2305. 2445

/rpt_family="MIR"

3035. 4004

/rpt_family="L1"

3997. 4042

/rpt_family="AT_rich"

4124. 4486

/rpt_family="L2"

4604. 4868

/rpt_family="L2"

6750. 7060

/rpt_family="Alu"

7036. 7060

/rpt_family="(A)n"

7919. 8009

/rpt_family="L2"

9556. 9594

/rpt_family="AT_rich"

9696. 9983

/rpt_family="L1"

9984. 10491

/rpt_family="L1"

10804. 10990

/rpt_family="MER1_type"

11032. 11148

/rpt_family="L2"

12179. 12264

/rpt_family="MER1_type"

13476. 13780

/rpt_family="Alu"

14016. 14059

/rpt_family="(TTTTTA)n"

14032. 14357

/rpt_family="Alu"

14218. 14241

/rpt_family="(TTTGTG)n"

14833. 15131

/rpt_family="Alu"

15112. 15140

/rpt_family="AT_rich"

15942. 16244

/rpt_family="Alu"

16250. 16276

/rpt_family="(TG)n"

16276. 16439

/rpt_family="(TA)n"

17191. 17680

/rpt_family="L1"

17443. 17493

/rpt_family="AT_rich"

17689. 17914

/rpt_family="L1"

17922. 18967

/rpt_family="L1"

18996. 19221

/rpt_family="L1"

19222. 19521

repeat_region /rpt_family="Alu"
19522. 19642
repeat_region /rpt_family="L1"
19657. 20071
repeat_region /rpt_family="MER1_type"
20111. 20591
repeat_region /rpt_family="L1"
20592. 21409
repeat_region /rpt_family="L1"
21410. 21738
repeat_region /rpt_family="L1"
22413. 22463
repeat_region /rpt_family="(A)n"
23739. 23838
repeat_region /rpt_family="ERV1"
23867. 23929
repeat_region /rpt_family="ERV1"
24308. 24751
repeat_region /rpt_family="ERV1"
24798. 24886
repeat_region /rpt_family="MER1_type"
26974. 27312
repeat_region /rpt_family="ERV1"
27313. 27722
repeat_region /rpt_family="ERV1"
27723. 27763
repeat_region /rpt_family="ERV1"
28013. 28120
repeat_region /rpt_family="ERV1"
28205. 28792
repeat_region /rpt_family="ERV1"
28800. 28941
repeat_region /rpt_family="MIR"
29026. 29069
repeat_region /rpt_family="(TG)n"
29554. 30100
repeat_region /rpt_family="AcHobo"
30098. 30628
repeat_region /rpt_family="AcHobo"
30840. 31241
repeat_region /rpt_family="L1"
31266. 31810
repeat_region /rpt_family="ERV1"
31869. 32241
repeat_region /rpt_family="MaLR"
32174. 32194

Query Match 5.3%; Score 61.6; DB 9; Length 169514;

Best Local Similarity 52.3%; Pred. No. 0.0011;

Matches 136; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 863 AAGTTCCTTCATACCTCTTTCTCTTTCCACTAGTCTGCTCTTTCT 922
Db 161422 AATAGCGTTTCTCAAACTTTTGGTCTTTGCACCCCTTACACTCTGAAAATTTATGAG 161363
Qy 923 TCTCTTAGATACCCCTCTTTTCAGGAGCTCTGCTCTACTATTGTGTCATTTCTCGAAAC 982
Db 161362 AACGAGAGAGAGCTTTCTTTTACTGCTTATCTTACTATATTTTACCATATCAGAAGC 161303
Qy 983 ATTCTCTCCCGTGCATTTTCCCTTTTATATATATATATATATATATATATATATATATAT 1042
Db 161302 TTACACAGACACTTTCTCAT 161243
Qy 1043 GTCCTCTTACGTATTTTGTATTTCTGTGCTTTTCAACAGATAGTCTATATATACGTTT 1102
Db 161242 TTATATATATATATTTTAT 161183
Qy 1103 GATACAGCTAGATATCGCTA 1122
Db 161182 TATACATTTATATATATTTA 161163

RESULT 14

AC129452/c	AC129452	138859 bp	DNA	linear	HTG 24-AUG-2002			
LOCUS	Rattus norvegicus clone CH230-918, *** SEQUENCING IN PROGRESS ***,							
DEFINITION	55 unordered pieces.							
ACCESSION	AC129452							
VERSION	AC129452.2	GI:22474826						
KEYWORDS	HTG; HTGS_PHASE1.							
SOURCE	Norway rat.							
ORGANISM	Rattus norvegicus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.							
AUTHORS	1 (bases 1 to 138859) Muzny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anquiano,D., Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,D., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mayhew,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A., Miner,G., Minjasa,M., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuon,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabot,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.							
TITLE	Direct Submission							
JOURNAL	Unpublished							
REFERENCE	2 (bases 1 to 138859)							
AUTHORS	Worley,K.C.							
TITLE	Direct Submission							
JOURNAL	Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA							
REFERENCE	3 (bases 1 to 138859)							
AUTHORS	Rat Genome Sequencing Consortium.							
TITLE	Direct Submission							
JOURNAL	Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA							
REFERENCE	On Aug 24, 2002 this sequence version replaced gi:22004139.							
AUTHORS	Center: Baylor College of Medicine							
TITLE	Center code: BCM							
JOURNAL	Web site: http://www.hgsc.bcm.tmc.edu/							
REFERENCE	Contact: hgsc-help@bcm.tmc.edu							
AUTHORS	----- Project Information							
TITLE	Center project name: GDEB							
JOURNAL	Center clone name: CH230-918							
REFERENCE	----- Summary Statistics							
AUTHORS	Sequencing vector: Plasmid;							
TITLE	Chemistry: Dye-terminator Big Dye; 100% of reads							
JOURNAL	Assembly program: Phrap; version 0.990329							
REFERENCE	Consensus quality: 81630 bases at least Q40							
AUTHORS	Consensus quality: 86480 bases at least Q30							
TITLE	Consensus quality: 90155 bases at least Q20							
JOURNAL	-----							
REFERENCE	* NOTE: Estimated insert size may differ from sequence length							
AUTHORS	* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).							
TITLE	* NOTE: This is a 'working draft' sequence. It currently							
JOURNAL	* consists of 55 contigs. The true order of the pieces							
REFERENCE	* is not known and their order in this sequence record is							
AUTHORS	* arbitrary. Gaps between the contigs are represented as							
TITLE	* runs of N, but the exact sizes of the gaps are unknown.							
JOURNAL	* This record will be updated with the finished sequence							
REFERENCE	* as soon as it is available and the accession number will							
AUTHORS	* be preserved.							
TITLE	1							
JOURNAL	1199: contig of 1199 bp in length							
REFERENCE	1200: gap of unknown length							
AUTHORS	1299: contig of 1319 bp in length							
TITLE	1300: contig of 1319 bp in length							
JOURNAL	2618: gap of unknown length							
REFERENCE	2718: contig of 1736 bp in length							
AUTHORS	4454: contig of 1736 bp in length							
TITLE	4554: gap of unknown length							
JOURNAL	5901: contig of 1347 bp in length							
REFERENCE	6001: gap of unknown length							
AUTHORS	7156: contig of 1155 bp in length							
TITLE	7256: gap of unknown length							
JOURNAL	8330: contig of 1074 bp in length							
REFERENCE	8430: gap of unknown length							
AUTHORS	8486: contig of 1056 bp in length							
TITLE	9586: gap of unknown length							
JOURNAL	10731: contig of 1145 bp in length							
REFERENCE	10831: gap of unknown length							
AUTHORS	11831: contig of 1000 bp in length							
TITLE	11832: gap of unknown length							
JOURNAL	11932: contig of 1318 bp in length							
REFERENCE	13250: gap of unknown length							
AUTHORS	13350: contig of 1188 bp in length							
TITLE	14537: gap of unknown length							
JOURNAL	14638: contig of 1543 bp in length							
REFERENCE	16180: gap of unknown length							
AUTHORS	16280: gap of unknown length							
TITLE	16281: contig of 1457 bp in length							
JOURNAL	17737: gap of unknown length							
REFERENCE	17837: contig of 1524 bp in length							
AUTHORS	19361: gap of unknown length							
TITLE	19461: gap of unknown length							
JOURNAL	20517: contig of 1056 bp in length							
REFERENCE	20617: gap of unknown length							
AUTHORS	21802: contig of 1185 bp in length							
TITLE	21902: gap of unknown length							
JOURNAL	23000: contig of 1098 bp in length							
REFERENCE	23100: gap of unknown length							
AUTHORS	24171: contig of 1071 bp in length							
TITLE	24271: gap of unknown length							
JOURNAL	25962: contig of 1691 bp in length							
REFERENCE	26062: gap of unknown length							
AUTHORS	27405: contig of 1343 bp in length							
TITLE	27505: gap of unknown length							
JOURNAL	28600: contig of 1095 bp in length							
REFERENCE	28700: gap of unknown length							
AUTHORS	30122: contig of 1422 bp in length							
TITLE	30222: gap of unknown length							
JOURNAL	30123							

COMMENT On Aug 21, 2002 this sequence version replaced gi:17060594.
Research, 320 Charles Street, Cambridge, MA 02141, USA

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

QY 1043 TCTACGGTATTTTGGTATTTCTGIGIGCTTTATCAAGGATAGCTATAAATACGTTTGATACA TTU08

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 05:47:57 ; Search time 312 Seconds
(without alignments)
8358.378 Million cell updates/sec

Title: US-09-814-661A-1
Perfect score: 1158
Sequence: 1 aatgagcaacgtgtcaacaa.....cctctctgtatcaatgcttt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0.

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

```

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1158	100.0	1158	21	Yeast SML1 (suppre
2	74	6.4	1767	17	Yeast calcineurin
3	57.6	5.0	5455	24	Chemically treated
4	57.6	5.0	5455	24	Human polynucleoti
5	54.2	4.7	394	23	Human prostate exp
6	54	4.7	4167	22	Human ovarian and
7	54	4.7	4167	22	Genomic sequence #
8	54	4.7	4167	22	Human lung antigen
9	54	4.7	4167	22	Human reproductive

10	54	4.7	4167	22	AA107543	Human reproductive
11	54	4.7	4167	22	AA28711	Genomic sequence #
12	54	4.7	4167	22	AAK68040	Human immune/haema
13	54	4.7	4167	22	AAK89382	Human digestive sy
14	54	4.7	4167	23	ABL97915	Human testicular a
15	53.6	4.6	6419	24	ABL32267	Human immune syste
16	53	4.6	509	24	ABQ16728	Oligonucleotide fo
17	53	4.6	509	24	ABQ16729	Oligonucleotide fo
18	53	4.6	220895	24	ABK84798	Human cDNA differe
19	52.6	4.5	5387	24	ABN80041	Human chemically m
20	52.4	4.5	24259	22	AA546691	Tumour suppressor
21	52.2	4.5	16033	24	ABL33404	Human immune syste
22	52	4.5	6301	24	ABL32053	Human immune syste
23	52	4.5	16602	24	ABN80068	Human chemically m
24	52	4.5	16602	24	ABL32726	Human immune syste
25	51.8	4.5	169739	24	ABQ88186	Human osteoblast d
26	51.4	4.4	408	23	ABV34679	Human prostate exp
27	51.4	4.4	408	23	ABV43535	Human prostate exp
28	51.4	4.4	6145	24	ABL32972	Human immune syste
29	51.2	4.4	556	23	ABV40063	Human prostate exp
30	51.2	4.4	556	23	ABV40163	Human prostate exp
31	51.2	4.4	556	23	ABV42105	Human prostate exp
32	51.2	4.4	556	23	ABV43601	Human prostate exp
33	51.2	4.4	8392	24	ABL33490	Human immune syste
34	50.8	4.4	6668	24	ABL33697	Human immune syste
35	50.8	4.4	14006	24	ABL33958	Human immune syste
36	50.6	4.4	598	24	ABQ52554	Oligonucleotide fo
37	50.6	4.4	598	24	ABQ52555	Oligonucleotide fo
38	50.6	4.4	600	24	ABQ52494	Oligonucleotide fo
39	50.6	4.4	600	24	ABQ52495	Oligonucleotide fo
40	50.6	4.4	8079	24	ABL92313	Chemically treated
41	50.6	4.4	16633	24	ABN79984	Human chemically m
42	50.4	4.4	693	24	ABQ32104	Oligonucleotide fo
43	50.4	4.4	693	24	ABQ32105	Oligonucleotide fo
44	50.4	4.4	5195	24	ABL32921	Human immune syste
45	50.4	4.4	6115	24	ABL336801	Human immune syste

ALIGNMENTS

RESULT 1
AA14263
ID AA14263 standard; DNA; 1158 BP.

AC AA14263;

DT 15-AUG-2000 (first entry)

DE Yeast SML1 (suppressor of mecl lethality) gene.

XX SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;
KW Yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mecl; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT; ds.

XX Saccharomyces cerevisiae.

XX Key Location/Qualifiers

FT CDS 549..863

FT /*tag= a

FT /product= "Yeast Sml1 protein"

FT /function= "Inhibitor of ribonucleotide reductase"

XX WO200017225-A2.

PN 30-MAR-2000.

PD 24-SEP-1999; 99WO-US22260.

XX 24-SEP-1998; 98US-0158858.

XX

XX PF 06-APR-2001; 2001WO-EP03971.
XX XX
XX OS 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX
XX PA (EPIC-) EPIGENOMICS AG.
XX XX
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-017471/02.
XX XX
XX PT New nucleic acid sequences from chemically modified genes associated
XX PT with DNA replication, useful for analysing cytosine methylations for
XX PT diagnosis and therapy of diseases e.g. Ataxia telangiectasia
XX XX
XX PS Claim 1; SEQ ID NO 33; 23pp + Sequence Listing; English.
XX XX
XX CC The invention relates to nucleic acid sequences comprising at least 18
XX CC bases of a chemically pretreated gene associated with gene regulation,
XX CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
XX CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPAL,
XX CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases
XX CC unmodified at the 5-position to uracil or another base with
XX CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
XX CC cytosine methylations. The DNA sequences and method are useful in the
XX CC diagnosis of diseases (or predisposition to diseases) associated with DNA
XX CC replication and in therapy of such diseases, by enabling analysis of the
XX CC cytosine methylation patterns of such genes. They are especially useful
XX CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
XX CC syndrome, solid tumours and cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification and was supposed to be available directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences. However, the sequence data did
XX CC not correspond to that referred to in the specification. The present data
XX CC is taken from EPO data for the patent.
XX XX
XX SQ Sequence 5455 BP; 1285 A; 149 C; 1404 G; 2617 T; 0 other;
Query Match 5.0%; Score 57.6; DB 24; Length 5455;
Best Local Similarity 54.8%; Pred. No. 0.00022;
Matches 114; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 856 TTCCTTTTCATCTCTTTCTTTCTTTCCATTTCCACACTAGTTCTGTTCTTTCTTCT 925
DB 665 TTCGATTATTTTGTAGTTTGTGTTTAAATTCGATTGTTTGTGTTGTTTCG 724
QY 926 CTTAGATACCTCTCTTTTCAGGACTCTGCGCTACTATTTGTCATTCGGAACATT 985
DB 725 TTTTAAAGTTTAAATTAATTTGTTTATTTTATTTTAGGTTTAAATATATATCGT 784
QY 986 CTCCTCCGTCGATTTCTTCCCTTTATATATATATATATATATATATATATATAT 1045
DB 785 TTTTGAAGCGTTTATTTATATATATATATATATATATATATATATATATATATA 844
QY 1046 TCTCTACGATTTTGTGTTCTGTT 1073
DB 845 TATTTTATTTATTTATTTATTTT 872
RESULT 5
ABV13564/C
ID ABV13564 standard; cDNA; 394 BP.
XX XX
XX AC ABV13564;
XX XX
XX DT 13-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 13555.
XX XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.
XX XX
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX XX
XX PD 23-AUG-2001.
XX XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX XX
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX XX
XX DR WPI; 2001-662795/76.
XX XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer
XX XX
XX PS Claim 1; Page 2248; 11750pp; English.
XX XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX XX
XX SQ Sequence 394 BP; 254 A; 38 C; 38 G; 61 T; 3 other;
Query Match 4.7%; Score 54.2; DB 23; Length 394;
Best Local Similarity 49.3%; Pred. No. 0.0006;
Matches 137; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 836 GGGTAAGTCGAGGAAATGGACTTCTAAAGTTCCTTCATACCTCTTTCTCTCTTC 895
DB 384 GGGGAAGTCCTCCGAGGAGAAAATTTTATTTTATTTTATTTTATTTTATTTT 325
QY 896 CATTTCCACTAGTCT 955
DB 324 TTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 265
QY 956 TCCTACTATGTTGTCATCTCTCGAACAATCTCTCCCGTGCATTTTCCCTTTCCCT 1015
DB 264 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 205
QY 1016 TACAT 1075
DB 204 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 145
QY 1076 TTATCAAGATAGTCTATATATACCTTTTCATACACTAG 1113
DB 144 TGAAGAATCACTCTTATATAACTTTAATTTATCTTG 107
RESULT 6

ABA08224
ID ABA08224 standard; DNA; 4167 BP.
XX
AC ABA08224;
XX
XX
DT 11-JAN-2002 (first entry)
XX
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 1019.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ds.
XX
OS Homo sapiens.
XX
XX
PN WO200155325-A2.
XX
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01345.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205513.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0246174.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.

DT 21-NOV-2001 (first entry)
XX Human lung antigen genomic DNA #88.
DE Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
XX chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
XX tissue regeneration; anti-infertility; food additive.
OS Homo sapiens.
XX WO200155303-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01301.
PF 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227189.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen -
is used in preventing, treating or ameliorating a medical condition -

Disclosure; SEQ ID NO 10231; 1297pp + Sequence Listing; English.

```

XX CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 4167 BP; 576 A; 990 C; 915 G; 686 T; 1000 other;

Query Match 4.7%; Score 54; DB 22; Length 4167;
Best Local Similarity 53.3%; Pred. No. 0.0016;
Matches 114; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 866 TTCTCTTACATCCTTTCTTTCTCTCTTTCCATTTCCCACTAGTTCGTCTTTTCTTCTCT 925
   || || || || || || || || || || || || || || || || || || || || ||
DB 44 TTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 103
   || || || || || || || || || || || || || || || || || || || || ||
QY 926 CTTAGATACCCCTTCTTTTTCAGGGACTCTCGTCTACTATTGTTGTCATTCGAAACATT 985
   || || || || || || || || || || || || || || || || || || || || ||
DB 104 TTTTTCCTCTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 163
   || || || || || || || || || || || || || || || || || || || || ||
QY 986 CTCCTCCGGGCATTTTCCTTTCCCTTTATATACATATATATATATATATATATATATGC 1045
   || || || || || || || || || || || || || || || || || || || || ||
DB 164 TTCTCTCTTTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 223
   || || || || || || || || || || || || || || || || || || || || ||
QY 1046 TCTCTACGTATTTTGTATTCTCTGTCTTTAT 1079
   || || || || || || || || || || || || || || || || || || || || ||
DB 224 TCTCTCTTTTTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 257
   || || || || || || || || || || || || || || || || || || || || ||

RESULT 11
AAS28711/C
ID AAS28711 standard; DNA; 4167 BP.
XX
AC AAS28711;
XX
XX
XX 07-NOV-2001 (first entry)
XX
XX Genomic sequence #551 encoding for novel human respiratory antigen.
XX
XX Human; respiratory antigen; respiratory disorder; throat disorder;
XX lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
XX anti allergic; anti asthmatic; anti inflammatory; oilfactory;
XX respiratory active; ds.
XX
XX Homo sapiens.
XX
XX WO200155448-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01333.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
XX
XX 14-AUG-2000; 2000US-0224519.

```


[illegible]


```
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 29-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
XX useful for preventing, diagnosing and/or treating testicular cancer
XX
XX Disclosure; SEQ ID NO 2567; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
XX human testicular antigens, and fragments of their genomic sequences. The
XX sequences can be used in the treatment of cardiovascular, urinary system,
XX reproductive system, immune, respiratory, neurological and
XX gastrointestinal disorders, infections, and particularly cancer,
XX especially testicular cancers. The present sequence is a DNA encoding a
XX protein fragment of the invention.
XX
XX Sequence 4167 BP; 576 A; 990 C; 915 G; 686 T; 1000 other;

Query Match 4.7%; Score 54; DB 23; Length 4167;
Best Local Similarity 53.3%; Pred. No. 0.0016;
Matches 114; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 866 TTCCTTTTCATCTCTTTCTCTTCCATFTCCACACTGTTCTGTTCTTTCTTCT 925
Db || || || || || || || || || || || || || || || || || || || ||
QY 926 CTTAGATACCTTCTTTTCAGGACTCTGCTCTACTATGTTGTCATTCGAAACATT 985
Db || || || || || || || || || || || || || || || || || || || ||
QY 104 TTTTCTCTCTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 163
Db || || || || || || || || || || || || || || || || || || || ||
QY 986 CTCTCCCGGCATTTTCTTCCCTTATATACATATATATATATATATATATATGTC 1045
Db || || || || || || || || || || || || || || || || || || || ||
QY 164 TTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 223
Db || || || || || || || || || || || || || || || || || || || ||
QY 1046 TCTTCTACGTATTTTGTATTTCTGTGCTTTAT 1079
Db || || || || || || || || || || || || || || || || || || || ||
QY 224 TCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 257
Db || || || || || || || || || || || || || || || || || || || ||

RESULT 15
ABL32267
ID ABL32267 standard; DNA; 6419 BP.
XX
XX ABL32267;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 240.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytosatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX Homo sapiens.
OS
```

XX WO200200928-A2.
XX
XX
PD 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
PR
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
XX Claim 1; SEQ ID NO 240; 32pp + Sequence Listing; German.
PS
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 6419 BP; 1565 A; 53 C; 1453 G; 3348 T; 0 other;
SQ

Query Match 4.6%; Score 53.6; DB 24; Length 6419;
Best Local Similarity 50.8%; Pred. No. 0.0024;
Matches 128; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 866 TTCCTTTCATCTCTTTCTCTTCCATTTCCACACTAGTCTCTCTTTCTTCT 925
DB || || | || || || || || || || || || || || || || || || || ||
1515 TTTATTTTATTTTATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTT 1574
QY 926 CTTAGATACCTCTTTTCAGGACTCTCGCTACTATTGTGTCATTCGAAACATT 985
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1575 TATTTTCTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1634
QY 986 CTCCTCCCGCATTTTCCTTTCCCTTTATATACATATATATATATATATATGTC 1045
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1635 TATTTATTTTATTTTATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1694
QY 1046 TCTTCTAGCTATTTTGTATTTCTGTCTTATCAAGATAGTCTATAACGTTTGCAT 1105
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1695 TTTATTTTATTTTGTATTTATTTTATTTTATTTTATTTTATTTTATTTATTT 1754
QY 1106 ACAGCTAGATAT 1117
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1755 TTATTTATTTAT 1766

Search completed: July 17, 2003, 11:49:56
Job time : 315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 11:44:36 ; Search time 85 Seconds
(without alignments)
4178.020 Million cell updates/sec

Title: US-09-814-661A-1
Perfect score: 1158
Sequence: 1 aatgagcaaccggtcaaca.....cctcttggatcaatgcttt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15333831 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfilesi.seq.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	6.4	1767	1	US-08-328-322-20
2	74	6.4	2353	1	US-08-328-322-11
3	74	6.4	7218	1	US-08-232-463-14
4	47.4	4.1	19124	2	US-08-487-8268-13
5	46.8	4.0	240	1	US-08-628-417-6
6	46	4.0	2447	2	US-09-014-969-14
7	46	4.0	6243	2	US-09-056-075-1
8	45.6	3.9	17949	4	US-09-087-465-3
9	45.4	3.9	2946	3	US-08-968-563-6
10	45.4	3.9	2946	3	US-08-969-683A-6
11	45.4	3.9	2946	4	US-09-297-928-2
12	44.8	3.9	162450	4	US-09-345-882-1
13	43.8	3.8	2674	4	US-09-019-095A-1
14	43.6	3.8	6124	4	US-08-213-419B-3
15	43.4	3.7	8920	2	US-08-446-855A-1
16	43.4	3.7	8920	4	US-09-150-741-1
17	43.2	3.7	2671	6	5168051-9
18	43	3.7	87350	3	US-08-781-891-79
19	43	3.7	87543	4	US-09-791-211-3
20	42.4	3.7	658	4	US-08-998-416-595
21	42.4	3.7	72604	4	US-09-268-992-7
22	42.4	3.7	72604	4	US-09-657-474-7
23	42.2	3.6	289	4	US-09-007-005-17
24	42.2	3.6	289	4	US-09-244-796-17
25	42.2	3.6	44453	4	US-09-146-053-5
26	42	3.6	176373	3	US-09-128-155-17
27	41.8	3.6	1142	1	US-08-006-082A-5

c	28	41.8	3.6	5394	3	US-08-688-376-1	Sequence 1, Appli
	29	41.6	3.6	5224	4	US-09-033-428-2	Sequence 2, Appli
	30	41.6	3.6	5224	4	US-09-033-556-5	Sequence 5, Appli
	31	41.6	3.6	152331	3	US-09-128-155-16	Sequence 16, Appli
	32	41.4	3.6	6152	4	US-08-973-462-1	Sequence 1, Appli
	33	41.2	3.6	43795	3	US-08-742-185-101	Sequence 101, App
	34	41	3.5	277	4	US-09-007-005-3	Sequence 3, Appli
	35	41	3.5	277	4	US-09-244-796-3	Sequence 3, Appli
	36	41	3.5	29604	3	US-08-781-891-207	Sequence 207, App
	37	40.8	3.5	2663	1	US-08-136-743B-3	Sequence 3, Appli
	38	40.8	3.5	5526	3	US-08-751-359-21	Sequence 21, Appli
	39	40.8	3.5	5526	4	US-08-907-146-21	Sequence 21, Appli
	40	40.8	3.5	84495	4	US-09-797-906-3	Sequence 3, Appli
	41	40.8	3.5	152331	3	US-09-128-155-16	Sequence 16, Appli
	42	40.4	3.5	1683	4	US-09-009-443-6	Sequence 6, Appli
	43	40.4	3.5	2143	4	US-08-630-915A-11	Sequence 11, Appli
	44	40.2	3.5	11485	4	US-09-410-464-9	Sequence 9, Appli
	45	40.2	3.5	19124	2	US-08-487-826B-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1

US-08-328-322-20

: Sequence 20, Application US/08328322

: Patent No. 5723436

: GENERAL INFORMATION:

: APPLICANT: Huang, Laiqiang

: APPLICANT: Cyert, Martha S.

: TITLE OF INVENTION: Calcineurin Interacting Protein Compositions

: TITLE OF INVENTION: and Methods

: NUMBER OF SEQUENCES: 23

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Dehlinger & Associates

: STREET: 350 Cambridge Avenue, Suite 250

: CITY: Palo Alto

: STATE: CA

: COUNTRY: USA

: ZIP: 94306

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/328,322

: FILING DATE: 24-OCT-1994

: CLASSIFICATION: 530

: ATTORNEY/AGENT INFORMATION:

: NAME: Sholtz, Charles K.

: REGISTRATION NUMBER: P38,615

: REFERENCE/DOCKET NUMBER: 8600-0151.10

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (415) 324-0880

: TELEFAX: (415) 324-0960

: INFORMATION FOR SEQ ID NO: 20:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1767 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: double

: TOPOLOGY: unknown

: MOLECULE TYPE: DNA

: HYPOTHETICAL: NO

: ANTI-SENSE: NO

: ORIGINAL SOURCE:

: INDIVIDUAL ISOLATE: DNA fragment containing

: INDIVIDUAL ISOLATE: CNA2deltaC coding sequence

: FEATURE:

: NAME/KEY: CDS

: LOCATION: 262..1767

: US-08-328-322-20

NAME: Glaister, Debra
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC 369-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-864-7620
TELEFAX: 650-845-6504
TELEX:

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: GPD2
US-08-969-683A-6

Query Match 3.9%; Score 45.4; DB 3; Length 2946;
Best Local Similarity 49.0%; Pred. No. 0.0077;
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 911 CTGTTCTTTCTCTCTAGATACCCCTCTTTTCAGGACACTCTCTCTACTATTGTTGT 970
DB 1054 CTGTCATTCTAGTATTTTTTTTTTTTTTTTGGTTTACTTTTTTTTCTTCTGCTTT 1113
QY 971 CATTCCTGAACATCTCTCCGTCATTTTCCCTTTTATATACATATATATAT 1030
DB 1114 TTTCTCTGTTACTTTTCTCTAGTTTCTTCCCTCACTAAGCTTTTCTCTGATTAT 1173
QY 1031 ATATATATATATGCTCTCTCTACGATTTTTTGTATTTCTGTCTTTATCAAGATAGTC 1090
DB 1174 CCTTGGTCTCTCTCTACTCTCTTAGATTTTTTTTTTATATATATATTTTAAAGTTTA 1233
QY 1091 TATAATAGTTTGATACAGCTAGATATCGTAGCCCAACATTTGCCCTCTCTTGATC 1150
DB 1234 TGTATTTGGTAGATTCAATCTCTTCCCTTCTCTTCTGCTGCCCTTCTTATC 1293
QY 1151 AATGCTT 1157
DB 1294 AATGCTT 1300

RESULT 11

US-09-297-928-2
Sequence 2, Application US/09297928
Patent No. 6358716

GENERAL INFORMATION:

APPLICANT: BULTHUIS, BEN A.
GATENBY, ANTHONY A.
HAYNIE, SHARON L.
HSU, AMY K.
LAREAU, RICHARD D.

TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
GLYCEROL BY RECOMBINANT
ORGANISMS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENECOR INTERNATIONAL, INC.

STREET: 4 CAMBRIDGE PLACE

CITY: ROCHESTER

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 14618

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MICROSOFT WINDOWS 95

SOFTWARE: MICROSOFT WORD VERSION 7.0A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/297,928

FILING DATE: 11-May-1999

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/03602
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-99B1-PI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
TELEX: 6717325

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2946 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-297-928-2

Query Match 3.9%; Score 45.4; DB 4; Length 2946;
Best Local Similarity 49.0%; Pred. No. 0.0077;
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 911 CTGTTCTTTCTCTCTAGATACCCCTCTTTTCAGGACACTCTCTCTACTATTGTTGT 970
DB 1054 CTGTCATTCTAGTATTTTTTTTTTTTTTTTGGTTTACTTTTTTTTCTTCTGCTTT 1113
QY 971 CATTCCTGAACATTTCTCTCCGTCATTTTCCCTTTTATATACATATATATATAT 1030
DB 1114 TTTCTCTGTTACTTTTCTCTAGTTTCTTCCCTCACTAAGCTTTTCTCTGATTAT 1173
QY 1031 ATATATATATATGCTCTCTCTACGATTTTTTGTATTTCTGTCTTTATCAAGATAGTC 1090
DB 1174 CCTTGGTCTCTCTCTACTCTCTTAGATTTTTTTTTTATATATATATTTTAAAGTTTA 1233
QY 1091 TATAATAGTTTGATACAGCTAGATATCGTAGCCCAACATTTGCCCTCTCTTGATC 1150
DB 1234 TGTATTTGGTAGATTCAATCTCTTCCCTTCTCTTCTGCTGCCCTTCTTATC 1293
QY 1151 AATGCTT 1157
DB 1294 AATGCTT 1300

RESULT 12

US-09-345-882-1

Sequence 1, Application US/09345882

Patent No. 6399373

GENERAL INFORMATION:

APPLICANT: Bougueleret, Lydie

TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
SEQUENCE 1, AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.

FILE REFERENCE: GENSET.031A

CURRENT APPLICATION NUMBER: US/09/345,882

PRIOR FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: US 60/091,315

PRIOR FILING DATE: 1998-06-30

PRIOR APPLICATION NUMBER: US 60/111,909

PRIOR FILING DATE: 1998-12-10

NUMBER OF SEQ ID NOS: 140

SOFTWARE: Patent.pm

SEQ ID NO 1

LENGTH: 162450

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: allele

LOCATION: 72794

OTHER INFORMATION: 5-124-273 : polymorphic base A or G

FEATURE:

NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele

LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
LOCATION: 103783..103828

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 13:35:06 ; Search time 312 Seconds
(without alignments)
7656.907 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 1158

Sequence: 1 aatgagcaaccgtgtcaaca.....cctcttctgatcaatgcttt 1158

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1158	100.0	1158	11	US-09-814-661A-1
2	59	5.1	1223197	15	US-10-027-632-179264
3	54	4.7	4167	10	US-09-764-878-282
4	54	4.7	4167	10	US-09-764-860-1145
5	54	4.7	4167	11	US-09-764-846-345
6	54	4.7	4167	12	US-09-764-891-7710
7	54	4.7	4167	12	US-09-764-891-10231
8	54	4.7	4167	15	US-10-091-483-345
9	54	4.7	4167	15	US-10-079-854-282
10	54	4.7	4167	15	US-10-074-095-1145
11	54	4.7	4167	15	US-10-205-428-1019
12	52.8	4.6	684973	11	US-09-263-959-1
13	52.6	4.5	1223197	15	US-10-027-632-179264
14	52	4.5	792	15	US-10-027-632-168583
15	52	4.5	792	15	US-10-027-632-168584
16	52	4.5	792	15	US-10-027-632-168585

17	52	4.5	792	15	US-10-027-632-168586	Sequence 168586,	
18	51.4	4.4	431	11	US-09-960-352-5558	Sequence 5558, Ap	
19	51.4	4.4	446	11	US-09-960-352-3400	Sequence 3400, Ap	
20	50.8	4.4	408	11	US-09-960-352-1221	Sequence 1221, Ap	
21	50.8	4.4	771	15	US-10-027-632-133833	Sequence 133833,	
22	50.8	4.4	771	15	US-10-027-632-133834	Sequence 133834,	
c	23	50.4	428	11	US-09-960-352-573	Sequence 573, App	
24	50.2	4.3	459	11	US-09-960-352-7666	Sequence 7666, Ap	
25	50	4.3	2994	15	US-10-027-632-115981	Sequence 115981,	
26	50	4.3	7657	15	US-10-239-676-185	Sequence 185, App	
27	49.8	4.3	451	11	US-09-960-352-10262	Sequence 10262, A	
28	49.8	4.3	506	15	US-10-027-632-184189	Sequence 184189,	
29	49.8	4.3	506	15	US-10-027-632-184190	Sequence 184190,	
30	49.8	4.3	602	15	US-10-027-632-279490	Sequence 279490,	
c	31	49.8	4.3	1267	15	US-10-001-843-45	Sequence 45, Appl
32	49.8	4.3	1344	15	US-10-027-632-122632	Sequence 122632,	
33	49.4	4.3	6944	15	US-10-172-086-112	Sequence 112, App	
34	49.2	4.2	516	11	US-09-960-352-5785	Sequence 5785, Ap	
35	49.2	4.2	529	11	US-09-983-965-2109	Sequence 2109, Ap	
36	49	4.2	633	15	US-10-027-632-49019	Sequence 49019, Ap	
37	49	4.2	633	15	US-10-027-632-58198	Sequence 58198, A	
38	49	4.2	633	15	US-10-027-632-59205	Sequence 59205, A	
c	39	48.8	4.2	268	11	US-09-960-352-11467	Sequence 11467, A
c	40	48.8	4.2	327	11	US-09-960-352-4630	Sequence 4630, Ap
c	41	48.8	4.2	1890	15	US-10-027-632-98111	Sequence 98111, A
42	48.8	4.2	2854	15	US-10-027-632-265939	Sequence 265939,	
c	43	48.6	4.2	650	15	US-10-027-632-236575	Sequence 236575,
c	44	48.6	4.2	650	15	US-10-027-632-236576	Sequence 236576,
45	48.4	4.2	344	11	US-09-960-352-1036	Sequence 1036, Ap	

ALIGNMENTS

RESULT 1

US-09-814-661A-1
; Sequence 1, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PT-US
; CURRENT APPLICATION NUMBER: US/09/814,661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent version 3.1
; SEQ ID NO 1
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: S. Cerevisiae
US-09-814-661A-1

Query Match 100.0%; Score 1158; DB 11; Length 1158;
Best Local Similarity 100.0%; Pred. No. 5.8e-279;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AATGACCAACCGTGTCAACAAGAGTGTCAAGACCGGCTACTTATTTCCCAAGATCACGT	60
DB	1	AATGACCAACCGTGTCAACAAGAGTGTCAAGACCGGCTACTTATTTCCCAAGATCACGT	60
QY	61	TCCTTCTGCCAACATCATTGCCGTGCGCGGTCTTCTGACATTTGTAAGAA	120
DB	61	TCCTTCTGCCAACATCATTGCCGTGCGCGGTCTTCTGACATTTGTAAGAA	120
QY	121	TACTTCCAACTAAGAGCATCTTCTTTTTTTTGTAGCCCAATGATAGGAAGAACAA	180
DB	121	TACTTCCAACTAAGAGCATCTTCTTTTTTTTGTAGCCCAATGATAGGAAGAACAA	180
QY	181	TAGATTATAAATACGTGACAAATATAGTAGATATGTTTTTATGACCTCGTACATAG	240
DB	181	TAGATTATAAATACGTGACAAATATAGTAGATATGTTTTTATGACCTCGTACATAG	240

:	: LOCATION: (2901)	n	equals	a,t,g,	or
:	: NAME/KEY: SITE				
:	: LOCATION: (2902)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2903)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2904)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2905)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2906)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2907)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2908)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2909)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2910)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2911)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2912)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2913)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2914)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2915)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2916)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: -LOCATION: (2917)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2918)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2919)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2920)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2921)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2922)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2923)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2924)	n	equals	a,t,g,	or
:	: OTHER INFORMATION:				
:	: NAME/KEY: SITE				
:	: LOCATION: (2925)	n	equals	a,t,g,	or


```
; NAME/KEY: SITE
; LOCATION: (2924)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2925)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2926)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2927)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2928)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2929)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2930)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2931)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2932)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2933)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2934)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2935)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2936)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2937)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2938)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2939)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2940)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2941)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2942)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2943)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2944)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2945)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2946)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2947)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

```
; LOCATION: (2948)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2949)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2950)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2951)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2952)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2953)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2954)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2955)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2956)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2957)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2958)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2959)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2960)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2961)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

Query Match 4.7%; Score 54; DB 10; Length 4167;

Best Local Similarity 53.3%; Pred. No. 0.0046;

Matches 114; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 866 TTCCCTTCATACCTCTTTCTTTCTTTCCACTTCCACTAGTCTGTCTTTCTTCT 925

Db 4124 TTTTCTTT 4065

QY 926 CTTAGATACCCCTCTTTTCAGGACTCTCGTCTACTATTGTGTCATTCTCGAACATT 985

Db 4064 TTTTCTCTCTCTTT 4005

QY 986 CTCCTCCGTCATTTTCCCTTATATACATATATATATATATATATATATATGTC 1045

Db 4004 TTCTCTCTTT 3945

QY 1046 TCTTCTACGATTTTGTGATTTCTGTGCTTTAT 1079

Db 3944 TCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3911

RESULT 5

US-09-764-846-345

; Sequence 345, Application US/09764846

; Patent No. US20020102638A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT212

; CURRENT APPLICATION NUMBER: US/09/764,846


```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (317)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (318)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (319)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (322)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (323)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (326)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (327)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

```
Query Match 4.78; Score 54; DB 11; Length 4167;
Best Local Similarity 53.38; Pred. No. 0.0046;
Matches 114; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 866 TTCCTTTCATACCTCTTTCTCTCTCTCTCCATTTCCACCTAGTCTGTTCTTTCTCT 925
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44 TTTCTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 926 CTTAGATACCTCTTTCTCAGGACTCTCGCTACTATTCTGCTATTCGGAACATT 985
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 104 TTTTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 163
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 986 CTCCTCCGTCGATTTCTTTCCCTTTATATACATATATATATATATATATATAT 1045
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 TTCCTCTCTTTTCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 223
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1046 TCCTTACGTAATTTTCTGTAATTCGTCTTTAT 1079
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 TCCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 6

US-09-764-891-7710

; Sequence 7710, Application US/09764891

; Publication No. US20030077808A1

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7710
; LENGTH: 4167
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (273)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (274)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (277)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (278)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (279)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (281)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (282)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (285)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (286)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (287)
```


RESULT 7
US-09-764-891-10231
; Sequence 10231, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10231
; LENGTH: 4167
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (273)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (274)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (277)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (278)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (279)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (281)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (282)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (285)
; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE
; LOCATION: (286)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (287)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (288)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (290)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (294)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (296)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (299)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (302)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (303)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (304)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (305)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (307)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

```

: LOCATION: (310)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (311)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (312)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (313)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (314)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (315)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (316)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (317)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (318)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (319)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (320)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (321)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (322)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (323)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (324)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (325)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (326)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (327)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (328)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE

```

```

Query Match
Best Local Similarity 4.7%; Score 54; DB 12; Length 4167;
Matches 114; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 866 TTCCCTTACATCTTTCTTTCTTTCCACTTCCACTAGTCTGTCTTTCTTCT 925
Db 44 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 103
QY 926 CTTAGTACCCCTTTCTTTTCCAGGACTCTCGTCTACTATGTGTGTCATCT 985
Db 104 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 163
QY 986 CTCTCCCGTCATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1045
Db 164 TTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 223

```

```

QY 1046 TCTTCTACGTATTTTGTATTTCTGTCTTTAT 1079
Db 224 TCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 257

RESULT 8
US-10-091-483-345
: Sequence 345, Application US/10091483
: Publication No. US20030049650A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT12C1
: CURRENT APPLICATION NUMBER: US/10/091,483
: CURRENT FILING DATE: 2002-03-07
: NUMBER OF SEQ ID NOS: 348
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 345
: LENGTH: 4167
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (268)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (269)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (270)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (271)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (272)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (273)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (274)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (275)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (276)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (277)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (278)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (279)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (280)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (281)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (282)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (283)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature

```

```
LOCATION: (284)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (285)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (286)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (287)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (288)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (289)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (290)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (291)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (292)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (293)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (294)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (295)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (296)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (297)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (298)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (299)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (300)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (301)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (302)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (303)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (304)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (305)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (306)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (307)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (308)
```

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (309)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (310)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (311)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (312)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (313)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (315)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (316)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (317)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (318)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (319)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (320)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (321)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (322)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (323)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (324)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (325)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (326)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (327)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (328)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
```

Query Match 4.7%; Score 54; DB 15; Length 4167;

Best Local Similarity 53.3%; Pred. No. 0.0046;

Matches 114; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 866 TTCCTTTCATACCTCTTTCTCTTTCCCAATTCCTCCACTAGTTCGTCTTTCTCTTCT 925

Db 44 TTTTCTT 103

Qy 926 CTTAGATACCCCTTCCTTTCAGGACTCTCGTCTCTACTATGTGTGTCATTTCTCGAAACATT 985

[illegible]

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 168584
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-168584

Query Match 4.5%; Score 52; DB 15; Length 792;
Best Local Similarity 52.8%; Pred. No. 0.0062;
Matches 112; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 866 TTCCTTTCACTCTTTTCTTTTCTTTCCACTTTCCACTAGTTCTGTTCTTTTCTTCT 925
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 475 TTCCTTCTCCCGCGCTCCCTCTCTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCC 534
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 926 CTTAGATACCTTCTTTTCAGGGACTCTCGTCTACTATTGTTGTCATTTCTCGAAACATT 985
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 535 CTTCTTTTCTCTTTCT 594
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 986 CTCCTCCGTCATTTTCTTTCCCTTTATATACATATATATATATATATATATATATGTC 1045
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 595 TCCCT 654
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1046 TCTTCTACGATTTTCTGTTCTGTTCTTT 1077
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 655 TTTCT 686
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: July 17, 2003, 15:39:50
Job time : 319 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 11:53:05 ; Search time 24 Seconds
(without alignments)
4002.466 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 2069
Sequence: 1 aatgagcaaccgtgtcaaca.....cctcttctgatcaatgcttt 1158

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09814661/runat_16072003_115258_14514/app_query_fasta_1.1351
-DB=SwissProt_40 -QFWT=fastan -SUFFIX=rsrp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pt0 -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09814661@cgn_1.1.28 @runat_16072003_115258_14514 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	537	26.0	104	1 YMF8_YEAST	Q04964 saccharomyc
2	95.5	4.6	164	1 YCE1_YEAST	P25571 saccharomyc
3	94.5	4.6	885	1 YD03_YEAST	Q06639 saccharomyc
4	93	4.5	1723	1 AIM1_HUMAN	O9y4k1 homo sapien
5	91	4.4	104	1 YGE1_YEAST	P53182 saccharomyc
6	90	4.3	902	1 SYG1_YEAST	P40528 saccharomyc
7	87	4.2	533	1 YADC_SCHPO	Q09837 schizosacch
8	86.5	4.2	1505	1 CUT1_HUMAN	P39880 homo sapien
9	86	4.2	1786	1 UVRA_CHLTR	O84337 chlamydia t
10	85	4.2	1711	1 CHD1_MOUSE	P40201 mus musculu
11	83	4.1	1709	1 CHD1_HUMAN	O14646 homo sapien
12	82.5	4.0	516	1 SMP3_YEAST	Q04174 saccharomyc
13	82.5	4.0	825	1 RCAL_YEAST	P40341 saccharomyc
14	81	3.9	474	1 MIG1_KLULA	P50898 kluyveromyc
15	81	3.9	626	1 PC11_YEAST	P39081 saccharomyc
16	81	3.9	985	1 AGLU_ASPNG	P56526 aspergillus
17	81	3.9	1395	1 CUT1_MOUSE	P53564 mus musculu
18	80	3.9	284	1 COX3_LEITA	P14546 leishmania

19	80	3.9	557	1	OCN2_MOUSE	Q920e8 mus musculu
20	80	3.9	3092	1	IRAL1_YEAST	P18963 saccharomyc
21	79.5	3.8	334	1	GP12_HUMAN	P47775 homo sapien
22	79.5	3.8	409	1	NU4M_CAEEL	P24892 caenorhabdi
23	79.5	3.8	642	1	YA2A_SCHPO	O09701 schizosacch
24	79.5	3.8	1787	1	UVRA_CHLMO	Q9pk60 chlamydia m
25	78.5	3.8	319	1	O2S2_HUMAN	O9nqn1 homo sapien
26	78.5	3.8	368	1	YBX5_SCHPO	Q10203 schizosacch
27	78	3.8	509	1	CCBS_MARPO	P36180 marchantia
28	78	3.8	799	1	YDDU_ECOLI	P76129 escherichia
29	77.5	3.7	334	1	GP12_MOUSE	P35412 mus musculu
30	77.5	3.7	349	1	SCA3_MOUSE	O35609 mus musculu
31	77.5	3.7	378	1	Y267_AQUAE	O66624 aquifex aeo
32	77.5	3.7	510	1	CX56_CHICK	P29415 gallus gall
33	77.5	3.8	821	1	YK62_CAEEL	P34341 caenorhabdi
34	77.5	3.7	905	1	TLR3_MOUSE	O99mb1 mus musculu
35	77	3.7	433	1	YBL4_YEAST	P98211 saccharomyc
36	76.5	3.7	199	1	YE65_MYCPN	P75318 mycoplasma
37	76.5	3.7	444	1	NU4M_LOCOMI	Q36424 locusta mig
38	76.5	3.7	634	1	YCX3_EUGGR	P31916 euglena gra
39	76.5	3.8	840	1	YHT1_YEAST	P38835 saccharomyc
40	76	3.7	131	1	IMMN_ECOLI	P09182 escherichia
41	76	3.7	144	1	NU6M_CAEEL	P24885 caenorhabdi
42	76	3.7	334	1	GP12_RAT	P30951 rattus norv
43	76	3.7	376	1	CYB_PLAFA	Q02768 plasmodium
44	76	3.7	459	1	NU4M_RHIUN	O96068 rhinoceros
45	76	3.7	501	1	S3A3_MOUSE	Q9d554 mus musculu

ALIGNMENTS

RESULT 1

YMF8_YEAST STANDARD; PRT; 104 AA.
AC Q04964;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Hypothetical 11.8 kDa protein in OGGL-CNA2 intergenic region.
GN YML058W OR YM9958.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Devlin K., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z46729; CAA86717.1; -
DR SGD: S0004523; YML058W.
KW Hypothetical protein.
SQ SEQUENCE 104 AA; 11834 MW; 02735C9178E431B2 CRC64;

Alignment Scores:
Pred. No.: 2.73e-45 Length: 104
Score: 537.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.95% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x YMF8_YEAST (1-104)

Qy	8	AACCGTGTCAACAAGAGTGTCAAGACCGCTACTTATTTCGCC--AAGATCACAGTTCCGT	64
Db	319	AsnAspLeuAsnIleLeuIleGluThrGlnPheAsnGlyProLeuHisAspLeuValAla	338
Qy	65	TCTGCCAACATCATTTGCCGTGCAAGCTCGCGCGCTTTCTGTGACATTGGTAAGAATACT	124
Db	339	SerArgAsn-----SerIleGluArgAsnSerGlyIleSerGlnIleLeuLysPhePro	356
Qy	125	TCCAACTAAGACATGCTTCCTTTTTTTTGTAGGCCCAATGATAGGAAGAACAAATAGA	184
Db	357	SerGln-----	358
Qy	185	TTATAAATACGTCAGAATATAGTAGATATGTTTTTATTGTTAGACCTCGTACATAGGAAT	244
Db	359	--SerIleThrGlnThrLeuIleAsnLysTyrLeuSer---ThrIleThrGluThrAsn	376
Qy	245	AATTGACGTTTTTTTGGCCCAACATTTGAAATTTTTTTTTTTTACCTCGCGCTGAGCCC	304
Db	377	Ser-----	384
Qy	305	AAACGGCTCCACTACCGCGCGCTCGCCATTTTGGGAAGATCATCCGTCCCAAAAGGA	364
Db	385	LysArgLeu---LeuProIleValIleGluLeuPheProSerAsnThrIleAsnLysPro	403
Qy	365	AATAGCCATAACATATCGTTACTGTGTTTGGAACATCGCCCGTTCCGCCGATTCGCGCTC	424
Db	404	AsnSerLysAspPheGluThrIle-----PheGlnValPheSerVal	417
Qy	425	ACGGGTATAAAAAGAGATCTTTTTTTTCTCGCTGCTGCCCTTCCATTTTAAATGCTTT	484
Db	418	ThrAsnAspGlnLeuLeuAsnLeuGlyPheIleThrLeu-----CysLeu	432
Qy	485	ATCTGCTCTTTGTGATCTTACGGTCT-----CAC	514
Db	433	LeuIleLeuPheGluSerLeuAsnSerThrValLeuIleProLeuArgAspGluHis	452
Qy	515	TAACTCTCTTCAACTG---CTCAATAATTTCCGGCTATGCAAAATTCACCAAGACTACT	570
Db	453	LeuGlnLeuPheAsnValLeuPheAsnTyrLeuProLeuLeuLysSerAsnLeuThrThr	472
Qy	571	TTTACGCTCAAAATCGCTGCC-----	591
Db	473	LeuArgPheGluIleGluLysArgSerMetCysAsnIleGluThrLeuArgPheIleSer	492
Qy	591	-----	591
Db	493	LeuTrpLysTyrTyrGlnPheValMetAspThrSerSerSerSerPheValIleAsp	512
Qy	592	-----AACACAAACAAAGCCCTT	609
Db	513	TyrAspGluAspMetHisMetalCysLeuLeuSerLeuAsnHisGluThrGlnAsnGln	532
Qy	610	CCACATTGCGTACCGTACCACTGGCGGAATTATA-----GAAGGGTGCCTTTGCCACCTTA	663
Db	533	SerHisIleLeu-----ThrTrpAsnPheIlePheLysAsnTyrCysTrpArgHisLeu	550
Qy	664	TGGCTGAGGTTCCTATGTGTCTACTCAAAACTCCATCG-----GCAGCTCCGCTTCG	717
Db	551	PheLeuGlyGlnLeuProLeuProLeuMetSerGluProPheThrAsnSerThrProIleIle	570
Qy	718	CCTCGCTCTTCTCAT-----TAGAATGTGGGAAAAGG	750
Db	571	AspProLeuLeuAsnAsnAspPheGluLeuIleAspPheGluValAsnLeuMetLysTyr	590
Qy	751	ATTTGGAGGAGAGACTCAACTCTATCGATCATGACATGAACAACAATAATTTGTTCTTG	810
Db	591	LeuGlnSerLysAspGlnGlnLeuSerIleAspLysIleIleGlnLeuIleLysLeuLeu	610
Qy	811	CGGAACATAAATCTA-----TGTTCAACAGGGTAAAGTCGAGGAATGCG	855
Db	611	LysAsnLysAsnIleGluValSerGlnGlyCysLeuThrThrProSerIleIle-AsnAs	630

Qy	856	ACTTCTAAAGTTCCTTTTCATAC-----TCCTTTTCTTTCTCTTTCCATTTCCCACTAGT	909
Db	630	nileMeCaspSerLeuLeuIleTyrArgAsnSerMetLeuTyr---LeuAsnPheTyrLeuLeu	649
Qy	910	TCTCTTTCTTTTCTCTCTTAGA---TACCCTCTTTTTCAGGGAC-----	950
Db	649	uLeuGlnPheGluThrLeuLysAsnTyrAlaLys-PheAsnGluLeuLeuGluAspPheL	669
Qy	951	-----TCTCGTCCTACTATTGTTCTTCATCTCTCGAACAATCTCTCCCGTCGATTTTC	1002
Db	669	euGluLeuSerArgGluThrLeuPheValPheSerAsnLeuAlaAsnIleLysPheA	689
Qy	1003	CTTTCCCTTATATACATATATATATATATATATATATCTCTCTCTACGTATTTTTC	1062
Db	689	laGlyHisGluPheThrPheIleAsnLysSerIleValValLeuGlnThrLeuValLeuM	709
Qy	1063	TATTCTGTGTCTTTATCAAGA	1085
Db	709	etLeuLeuAlaLeuTyrGlnArg	716
RESULT 4			
A1M1_HUMAN			
ID	A1M1_HUMAN	STANDARD:	PRT: 1723 AA.
AC	Q9Y4KI; O00296;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Absent in melanoma 1 protein.		
GN	A1M1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	TISSUE=Liver;		
RX	MEDLINE=97250519; PubMed=9096375;		
RA	Ray M.E., Wistow G., Su Y.A., Meltzer P.S., Trent J.M.;		
RT	"A1M1", a novel non-lens member of the betagamma-crystallin		
RT	superfamily, is associated with the control of tumorigenicity in human		
RT	malignant melanoma";		
RL	Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).		
CC	!- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY		
CC	EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.		
CC	!- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS		
CC	6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.		
CC	!- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U83116; AAB53792.1; -;		
DR	EMBL; U83115; AAB53791.1; -;		
DR	HSSP; P02526; LGAM.		
DR	Genew; HGNC:356; A1M1.		
DR	MIN; 601797; -;		
DR	InterPro; IPR001064; Crystallin.		
DR	InterPro; IPR000772; Ricin_B_lectin.		
DR	Pfam; PF00030; crystall; 6.		
DR	Pfam; PF00652; Ricin_B_lectin; 3.		
DR	PRINTS; PR01367; BGCRCYSTALLIN.		
DR	SMART; SM00458; RICIN; 1.		
DR	SMART; SM00247; XTALbg; 6.		
DR	PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 7.		
DR	PROSITE; PS00231; RICIN_B_LECTIN; 1.		
KW	Repeat; Lectin.		
FT	DOMAIN 1022		
FT	DOMAIN 1123		
FT	DOMAIN 1206		
FT	CRYSTALLIN BETA/GAMMA-LIKE 1.		
FT	CRYSTALLIN BETA/GAMMA-LIKE 2.		

FT	DOMAIN	1207	1318	CRYSTALLIN BETA/GAMMA-LIKE 3.
FT	DOMAIN	1319	1404	CRYSTALLIN BETA/GAMMA-LIKE 4.
FT	DOMAIN	1405	1497	CRYSTALLIN BETA/GAMMA-LIKE 5.
FT	DOMAIN	1502	1584	CRYSTALLIN BETA/GAMMA-LIKE 6.
FT	DOMAIN	1586	1719	RICIN B-TYPE LECTIN.
SQ	SEQUENCE	1723 AA;	188646 MW;	7E50F681A627FB09 CRC64;

Alignment Scores:

Alignment Scores:	
Pred. No.:	0.526
Score:	93.00
Percent Similarity:	35.05%
Best Local Similarity:	25.43%
Query Match:	4.49%
DB:	1
Length:	1723
Matches:	78
Conservative:	24
Mismatches:	108
Indels:	81
Gaps:	13

US-09-814-661A-1 (1-1158) x AIM1_HUMAN (1-1723)

Qy	293	CGGCTGAGCCCAACAGGGGCTCGACCTACCCGCCCGCGGTGCCCACTATTTTGGGAAGTCATCCG	352
Db	778	LeuLeuArgProGlyLysArgAlaLaserAlaGluGlnSerValLeuPheLysSerLeuHisThr	797
Qy	353	TCCCAAAAGGAATAGCCATAACATATCTGTTGTTGGACATCGCCCGTTTCGCC	412
Db	798	AsnThrAsnGlyAsnSerGluProLeuValMetProGluLeuAsn-AspLysGluAsnAr	817
Qy	413	CGATTCCGCTACGCGGTATAAAAGAGAT-----CTTTTTTTTTTC	454
Db	817	gAspValThrAsnGlyGlyIleLysArgSerArgLeuGluLysSerAlaLeuPheSerSe	837
Qy	455	CTGGCTG---TCCCTCCATTTTAAATGCTTATCTGCTCCTTTGTGATCTCTACGGTCT	511
Db	837	rLeuLeuSerLeuLeuProGlnAspLysIlePheSerProSerValThrSer-----	854
Qy	512	CACATAACCTCTCTCAACTGCTCAATAAATTTCCGCTATGCAAAATTCACCAAGACTACTT	571
Db	855	-ValAsnThrMetThrThr-----AlaPheSerThrSerGlnAsnGlySe	869
Qy	572	TTACGCTCAAAATCGCTGCCAACAACAAGCCCTTCACATGTGCGTACCGTGACCAT	631
Db	869	rLeuSerGlnSerSerValSerGln-----ProThrThr-----	880
Qy	632	GGCGGAATTAGGAAGGGTGCCTTTGCCACCTATGGCT-----CAGGTTCTCTATGTT	682
Db	881	-----GluGlyAlaProProCysGlyLeuAsnLysGluGlnSerAsnLe	895
Qy	683	GTCTACTCAAAACTCCATGGGC-----AGCTCGCTTCCTGCCCTCCGCTCTCTC	730
Db	895	uLeuProAspAsnSerLeuLysValPheAsnPheAsnSerSerThrSerHisSerSe	915
Qy	731	ATTA-----CAAATGTGGAAAGGATTT	754
Db	915	rLeuLysSerProSerHisMetGluLysTyrProGlnLysGluLysThrLysGluAspLe	935
Qy	755	GGAGGAGAGACTCAAC-----TCAT	775
Db	935	uAspSerArgSerAsnLeuHisLeuProGluThrLysPheSerGluLeuSerLysLeuLy	955
Qy	776	CGATCATGACATGAACAACAACAAATTTGGTCTGCGAACAATAAAATCTATGTTCAACCA	835
Db	955	sAsnAspMetGluLysAlaAsnHisIleGluSerValIleLysSerAsnLeuProAs	975
Qy	836	GGTAAGTCGAGGAATGACATCTCTAAAGTTCCTTTTCATACCTCTTTCTCTCTCTTC	895
Db	975	nCysAlaAsnSerAspThrAspPheMetGlyLeuPheLysSerSerArgTyrAspProSe	995
Qy	896	CATTTCCTCCACTAGTCTGTTCTCTTTCTCTCTCTAGATACCTCTCTTTTCAGGACTCT--	953
Db	995	rIleSer-----PheSerGlyMetSerLeuSerAspThrMetThrLeuArgGlySerVa	1013
Qy	954	-----CGTCTACTATTTGTTGCATT-----	974
Db	1013	lGlnAsnLysLeuAsnProArgProGlyLysValIleLysSerGluProAspfalse	1033

QY	975	-----CTCGAAACATTCTCTCCGCG 995
		::: :::
Db	1033	rGluLysCysilecluValPheSerAspile 1043
RESULT 5		
YGEI_YEAST		
ID	YGEI_YEAST	STANDARD; PRT; 104 AA.
AC	P53182;	
DT	01-OCT-1996	(Rel. 34, Created)
DT	01-OCT-1996	(Rel. 34, Last sequence update)
DT	01-NOV-1997	(Rel. 35, Last annotation update)
DE	Hypothetical 12.0 kDa protein in DST1-HEM2 intergenic region.	
GN	YGL041C.	
OS	Saccharomyces cerevisiae (Baker's yeast).	
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	
OX	NCBI_TaxID=4932;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=S288C;	
RA	MEDLINE=97377993; PubMed=9234674;	
FX	Feuermann M., de Montigny J., Potier S., Souciet J.-L.;	
RT	"The characterization of two new clusters of duplicated genes	
RT	suggests a 'lego' organization of the yeast Saccharomyces cerevisiae	
RT	chromosomes."	
RL	Yeast 13:861-869(1997).	
RN	[2]	
RP	SEQUENCE OF 18-104 FROM N.A.	
RA	Hedling U., Hofmann B., Dellus H.;	
RL	Submitted (May/1996) to the EMBL/GenBank/DBJ databases.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; Z72563; CAA96743.1; -	
DR	SGD; S0003009; YGL041C.	
KW	Hypothetical protein; Transmembrane.	
FT	TRANSMEM 19 39	POTENTIAL.
FT	TRANSMEM 68 88	POTENTIAL.
SQ	SEQUENCE 104 AA; 11977 MW; 85A3B9FC9AE7AD93 CRC64;	
Alignment Scores:		
Pred. NO.:	0.56	Length: 104
Score:	91.00	Matches: 26
Percent Similarity:	46.74%	Conservative: 17
Best Local Similarity:	28.26%	Mismatches: 17
Query Match:	4.40%	Indels: 32
DB:	1	Gaps: 4
US-09-814-661A-1 (1-1158) x YGEI_YEAST (1-104)		
QY	871	TTGATACCTCTTTCTTTCTTTCTTTCCACTAGTCTGTTCTTTCTTTCTCTCTAG 930
Db	30	PheIleValPheSerLeuAlaGluValProSerArgLeuThrAsnPhePheSer--- 48
QY	931	ATACCCCTCTTTTCAGGAGCTCTCGTCTACTATTGTTGTCTTCGAAACATTCT--- 987
Db	49	-----IleMetIleLeuLeuThrPheSerAsnPheSerGln 60
QY	988	-----CTCCCGTGCATTTTCCTTTCCCTTTATATACATATATATATATATA 1041
Db	61	AsnIleArgProArgIleTyrLeu-----IleHisGluPheLeuHisLeuTyrVal 77
QY	1042	TGCTCTCTACGATTTTGTATTCTGTCTTTTCAAGTAGTCTTATATAGCTT 1101
Db	78	CysIleTyrPheValIleArg----- 84
QY	1102	TGATACAGCTAGATATACGCTAGGCCCAACATTGTCC 1137

```

Db 85 -----LeuSerValProArgLeuSer 92
RESULT 6
SYGL_YEAST STANDARD; PRT; 902 AA.
AC P40528; P40964;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE SVGL1 protein.
GN SVGL1 OR YIL047C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP1;
RA Spain B.H., Koo D., Ramakrishnan M., Dzduszor B., Colicelli J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowmen S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z46861; CAAG6904.1; -
CC ENBL: U14726; AAA91621.1; -
CC PIR: S48245; S48245.
CC SGD: S0001309; SYGL.
CC InterPro: IPR004342; EXS_Cterm.
CC InterPro: IPR004331; SPX.
CC Pfam: PF03105; SPX; 1.
CC Pfam: PF03124; EXS; 1.
CC Transmembrane.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 436 456 POTENTIAL.
FT TRANSMEM 498 518 POTENTIAL.
FT TRANSMEM 524 544 POTENTIAL.
FT TRANSMEM 555 575 POTENTIAL.
FT TRANSMEM 576 596 POTENTIAL.
FT TRANSMEM 674 694 POTENTIAL.
FT TRANSMEM 733 753 POTENTIAL.
FT TRANSMEM 762 782 POTENTIAL.
FT TRANSMEM 177 177 S -> T (IN REF. 1).
SQ SEQUENCE 902 AA; 104217 MW; F8D87D1DB3AED64 CRC64;

Alignment Scores:
Pred. No.: 0.954 Length: 902
Score: 90.00 Matches: 62
Percent Similarity: 37.8% Conservative: 46
Best Local Similarity: 21.75% Mismatches: 91
Query Match: 4.35% Indels: 86
DB: 1 Gaps: 15

US-09-814-661A-1 (1-1158) x SYGL_YEAST (1-902)
QY 234 TACATAGGAATAATTGACGTTTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTC 293

```


DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD00006; ABC_transport; 1.
 DR TIGRFAMs: TIGR00630; uvra; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
 KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
 KW DNA-binding; Zinc-finger; Complete proteome.
 FT NP_BIND 32 39 ATP (POTENTIAL).
 FT NP_BIND 625 632 ATP (POTENTIAL).
 FT NP_BIND 964 971 ATP (POTENTIAL).
 FT NP_BIND 1516 1523 ATP (POTENTIAL).
 FT ZN_FING 719 742 C4-TYPE.
 FT ZN_FING 1602 1628 C4-TYPE.
 SQ SEQUENCE 1786 AA; 196948 MW; 02D6862BE15DE070 CRC64;

Alignment Scores:
 Pred. No.: 2 62 Length: 1786
 Score: 86.00 Matches: 86
 Percent Similarity: 34.52% Conservative: 50
 Best Local Similarity: 21.83% Mismatches: 122
 Query Match: 4.16% Indels: 136
 DB: 1 Gaps: 18

US-09-814-661A-1 (1-1158) x UVRA_CHLTR (1-1786)

QY 227 GACCTGTCATAGGAATAAT-----TGACGTTTTTTTTGGCCCAACATT 271
 Db 899 GluLeuIleHisLysHisThrProThrAlaIleAlaLeuArgProPheLeuSerPro 918
 QY 272 TGAATTTTTTTTGTACTCGCGCTGAGCCCAACGGCTCCACTACCGCCGGTTC 331
 Db 919 GluLeuProTyrLeuProAspProSerProLys---ProProValProAlaIle 937
 QY 332 GCCATTGTTGGAGTCATCGCTCCCAAAAGGAATAGCCATAACATATCGTTACTGTT 391
 Db 938 ThrIleAlaAsnAlaHis-----GlnHisAsnLeuLysHisIleAsp 951
 QY 392 TGAACATCGCCGTTTCCCGATTCGCC-----TCAGCGGGTATAAAAG 439
 Db 952 LeuSerIleProArgTyrAlaLeu-ThrAlaValThrGlyProSerAlaSerGlyLysH 971
 QY 440 AGATCTTTTTTTTCTGCTGCTGCTCCCTCCATTTTAAATGCTTATCTGCTCTTGG 499
 Db 971 sserLeuValPhe-----AspIleLeuHisAla----- 980
 QY 500 ATCTTACGCTCTCACTAACCTCTTCTCAACTGCTCAATAATTTCCGCGTATGCAAAATTC 559
 Db 980 ----- 980
 QY 560 CCAAGACTACTTTACGCTCAAAATCGCTGCCAACACAAAGCCCTCCACATGCG 619
 Db 981 -----AlaGlyAsnIleAlaTyrAlaGluLeuPheProTyrIleAr 995
 QY 620 TACCGTGACCATGGCGGAATTTAGAGGCTGCTTTGCCACCTATGCGTGAGTT----- 674
 Db 995 g-----GlnAlaLeuIleLysHisThrProLeuProAlaValAspLysValThrG 1012
 QY 675 -----CCTATGTTGCTTACTCAAACTCCATGGGCAGC-----TCGCGTCTCGCTTC 721
 Db 1012 yLeuSerProValIleAlaIleGluLysThrSerAlaSerArgAsnSerAsnHisSerVa 1032
 QY 722 CGCTTCTTATAGAAATGTGGAAAGGATTTGGAGGAGAGACTCAACTCTATCATCA 781
 Db 1032 lAlaSerAlaLeuGluIleSerGluMet---LeuGluSerLeuPheThrArgIleGlyH 1051
 QY 782 TGACATGAACAACAATAATTTGGTTCTGCGCACTAAATCTATGTTCAACGAGGTAA 841
 Db 1051 sProTyr-----SerProIleSerGlyAspThrLeuArgThrIleSerProGluThr 1069
 QY 842 GTGTGAGGAA-----ATGGACTTCTAAAGTTCTCTTC- 873
 Db 1069 eAlaGluGluLeuLeuThrHisThrLysGlyTyrVal-ThrIleThrValProPheP 1089
 QY 874 -----ATACTCTTTCTCTTT- 889

Db 1089 rOLysGluGluGluPhePheSerTyrThrGlnGluMetLeuGlnGluGlyPheLeuLysL 1109
 QY 890 TCTTT-----CCATTTCCCACTAGT-----T 910
 Db 1109 euTyrAlaAsnGluGlnPheTyrAspLeuAspGlyProPheProThrSerLeuGluAsn 1129
 QY 911 CTTGTTCTTTCTCTCTTAGATACCTCTTTTCAGGGACTCTCGTCCCTACTATTTC- 966
 Db 1129 roAlaLeuValIleHisHisValLysIleLeuGluLysAsnLeuProSerLeuLeuAla 1149
 QY 967 -----TTGTCATTCTCGAAACATCTCTCCCGTGCATTTTTCCT- 1005
 Db 1149 erLeuThrLeuAlaPheSerLysAlaSerSerValCysLeuHisIleGluTyrAlaGly 1169
 QY 1005 ----- 1005
 Db 1169 hrSerLeuSerLysThrTyrArgGlnGlyLeuGlnAspAlaSerGlyAsnLeuPhePro 1189
 QY 1006 -----TCCCTTTATATACATATATATATATATATATATATATATATATATAT 1051
 Db 1189 snIleGluThrProSerIleLeuAsnHisGluSerTyrLeuCysProLeuCysHisGly 1209
 QY 1052 ACGTATTTTGTATTCTGCTGCTTTATCAAGATAGTCTATAATACGTTTGCATCAGCT 1111
 Db 1209 ysGlyPheLeuSerThrCysSerIleLeuProHisLysLysArgPheAlaGlnHisThr 1229
 QY 1112 AGATATCGCTAGCGCAACATTGTCCTCCCTCTCTTGCAT 1149
 Db 1229 rolleSerLeuPheThrSerLeuPheProAsnGlnAsp 1241

RESULT 10
 CHDL_MOUSE
 ID CHDL_MOUSE STANDARD; PRT; 1711 AA.
 AC P40201;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
 GN CHD1 OR CHD-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93211972; PubMed=8460153;
 RA Delmas V., Stokes D.G., Perry R.P.;
 RT "A mammalian DNA-binding protein that contains a chromodomain and an
 RT SNF2/SWI2-like helicase domain."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2414-2418(1993).
 CC -1- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
 CC -1- IMPORTANT ROLE IN GENE REGULATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: ABUNDANCE IS HIGHER IN CELLS REPRESENTING
 CC EARLY STAGES OF THE B LYMPHOID LINEAGE SUCH AS PRE-B AND B CELLS,
 CC THAN IN CELLS REPRESENTING MATURE PLASMACYTES OR OTHER CELL
 CC LINEAGES SUCH AS FIBROBLASTS.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 CC EMBL: L10410; AAB08486.1;
 CC PIR: A47392; A47392.
 CC HSSP: P23197; IAP0.
 CC MGD; MGI:88393; Chd1.

```

DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00176; SNF2_N; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00385; Chromo; 2.
DR SMART: SM00298; Chromo; 2.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00598; CHROMO_1; 2.
DR PROSITE: PS00013; CHROMO_2; 2.
KW DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT DOMAIN 1 70 SER-RICH.
FT DOMAIN 116 136 SER-RICH.
FT DOMAIN 270 362 CHROMO 1.
FT DOMAIN 387 450 CHROMO 2.
FT NP_BIND 504 511 ATP (POTENTIAL).
FT SITE 612 615 DEAH BOX.
FT DOMAIN 1629 1645 3 X 5 AA REPEATS OF H-S-D-H-R.
FT REPEAT 1629 1633 1.
FT REPEAT 1635 1639 2.
FT REPEAT 1641 1645 3.
SQ SEQUENCE 1711 AA; 196409 MW; FE3F8F8D13E32E24 CRC64;

Alignment Scores:
Pred. No.: 3.28 Length: 1711
Score: 85.00 Matches: 49
Percent Similarity: 36.90% Conservative: 44
Best Local Similarity: 19.44% Mismatches: 79
Query Match: 4.17% Indels: 80
DB: 1 Gaps: 6

US-09-814-661a-1 (1-1158) x CHD1_MOUSE (1-1711)
QY 896 GGAAGAGAAAGAAAGAGTATG-----AAAGACTTTAGAGTCCAT 852
DB 900 GlyGlnLysGlnValAsnIleTyArgLeuValThrLysGlySerValGluGluAsp 919
QY 851 TTCTCGACCTTACCCTGGTGAACATAGATTTTGTAGTCCGCAGAACCAATTTGTT 792
DB 920 IleLeuGluArg-----AlaLysLysMetValLeu 930
QY 791 GTTCATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 732
DB 931 AspHisLeuValIleGlnArgMetAspThr-ThrGlyLysThrValLeuHisThrGly 950
QY 731 T----- 731
DB 950 rAlaProSerSerThrPropheAsnLysGluGluLeuSerAlaIleLeuLysPheG1 970
QY 730 -----GAAGACGGAGCGAGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 687
DB 970 yAlaGluGluLeuPheLysGluProGluGlyGluGluGluGluGluGluGluGlu 990
QY 686 AGACACATAGAACCTCAGCCATAGGTGGCAAGGACCCCTCTAAATTCGCCCATGGT 627
DB 990 eAspGluLeuLeuLysArgAla-----G1 998
QY 626 CAGGTACGCAATGTGAAGGGCTGTGTTGTTGGCAGCGATTTTGACCGTAAAGTA 567
DB 998 uThrHisGluAsnGluProGlyPro-LeuSerValGlyAspGluLeuSerGlnPheL 1018
QY 566 GTCTTGGGAATTTGCATACCGG-----GAAATATTGACAGCTGAAGA 522
DB 1018 ysValAlaAsnPheSerAsnMetAspGluAspAspIleGluLeuGluProGluArgAsnS 1038
QY 521 GAGGTAGTGAGACCGCTAAGATCACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 462
DB 1038 erLysAsnTPGluGluIleProGluGluGlnArgArgGluLeuGluGluGluGlu 1058
QY 461 CAGCAGGAAAAAAGATCTCTTTTATACCGGTGAGCGGGAATCGCGGAAACGGG 402

```

```

Db 1058 rgGlnLysGluLeuGluGluIleTyArgMetLeuProArgMetArgAsnCysAlaLysGlnI 1078
QY 401 CGATGTTCCAAACAGTAAACGATATGTTATGGCTATTTTCCTTTTGGGACGGATGACTTC 342
Db 1078 leSerPheAsnGlySer----- 1083
QY 341 CCAAAATGGCGACCGCGGGTAGTGGAGCCGCTTTGGCTCAGCGGAGGTAACAAAA 282
Db 1084 -----GluGlyArgArgSerArgSerArgArgTySerGlyS 1096
QY 281 AAAAATTTCAATGTTGGCCCAAAAAAAGCGT 250
Db 1096 erAspSerAspSerIleSerGluArgLysArg 1106

RESULT 11
CHD1_HUMAN
ID CHD1_HUMAN STANDARD; PRT; 1709 AA.
AC O14646;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
GN CHD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97470991; PubMed-9326634;
RT Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT "Characterization of the CHD family of proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC -!- IMPORTANT ROLE IN GENE REGULATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EXBL; AF006513; AAB87381.1;
DR HSSP; P23197; IAP0.
DR Genew; HGNC:1915; CHD1.
DR MIM; 602118;
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00176; SNF2_N; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00385; Chromo; 2.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS00013; CHROMO_2; 2.
KW DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT DOMAIN 1 70 SER-RICH.
FT DOMAIN 117 137 SER-RICH.
FT DOMAIN 272 364 CHROMO 1.
FT DOMAIN 389 452 CHROMO 2.
FT NP_BIND 506 513 ATP (POTENTIAL).
FT SITE 614 617 DEAH BOX.
FT DOMAIN 1628 1644 3 X 5 AA REPEATS OF H-S-D-H-R.
FT REPEAT 1628 1632 1.
FT REPEAT 1634 1638 2.

```

```
FT REPEAT 1640 1644 3.
SQ SEQUENCE 1709 AA; 196517 MW; 416409C913D6A935 CRC64;

Alignment Scores:
Pred. No.: 5.18 Length: 1709
Score: 83.00 Matches: 42
Percent Similarity: 39.61% Conservative: 40
Best Local Similarity: 20.29% Mismatches: 67
Query Match: 4.07% Indels: 58
DB: 1 Gaps: 5

US-09-814-661A-1 (1-1158) x CHD1_HUMAN (1-1709)
QY 896 GGAAGAGAGAAAGAGAGTATG-----AAAGAACTTTAGAGTCCAT 852
Db 902 GlyGlnLysGlnValAsnIleTyrArgLeuValThrLysGlySerValGluGluAsp 921
QY 851 TTCTCGACCTTACCTGCTGTTGAACATAGATTTAGTCGCCAGAACCAAAATGTTGT 792
Db 922 IleLeuGluArg-----AlaLysLysLysMetValLeu 932
QY 791 GTTCATGTCATGATGATAGAGTGTGAGTCTCTCTCCCAAAATCCTTTTCCACATTTCTAA 732
Db 933 AspHisLeuValIleGlnArgMetAspThr-ThrGlyLysThrValLeuHisThrGlySe 952
QY 731 T----- 731
Db 952 rAlaProSerSerThrProPheAsnLysGluGluLeuSerAlaIleLeuLysPheG 972
QY 730 -----GAAGAGCGGAGCGAGAGCGGAGCGAGTGCCTGAGTGGATTTTGAGT 687
Db 972 yAlaGluGluLeuPheLysGluProGlyGluGluGluGluGluGluMetAspI 992
QY 686 AGACAACATAGGAACCTCAGCCATAGGTGCGCAAGGCCCTTCTAAATTCGCCCATGCT 627
Db 992 eAspGluIleLeuLysArgAla-----G 1000
QY 626 CACGTTACCAATGTGTGAAGGGCGTGTGTTGTTGGCAGCGATTTTGAGCGTAAAGTA 567
Db 1000 uThrHisGluAsnGluProGlyPro-LeuThrValGlyAspGluLeuSerGlnPheL 1020
QY 566 GTCTCGGGAATTTTGCATAGCGG-----CAAAATTTAGCAGATTTGAAGA 522
Db 1020 yValAlaAsnPheSerAsnMetAspGluAspIleGluLeuGluProGluArgAsn 1040
QY 521 GAGGTTAGTGAGACCGTAAGATCACAAAGGAGCAGATAGACATTTAAATGGAAGGA 462
Db 1040 erLysAsnTrpGluGluIleIleProGluAspGlnArgArgLeuGluGluGluGlu 1060
QY 461 CAGCCAGGAAAAAAGATCTCTTTTATACCGCTGAGCGGAATCGGCGAAACGGG 402
Db 1060 rgGlnLysGluLeuGluGluIleTyrMetLeuProArgMetArgAsnCysAlaLysGlnI 1080
QY 401 CGATGTTCCAAACAGCT 385
Db 1080 leSerPheAsnGlySer 1085

RESULT 12
SWP3_YEAST
ID SWP3_YEAST STANDARD; PRT; 516 AA.
AC Q04174; Q99400;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SMP3 protein.
GN SMP3 OR YOR149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NBW5;
```

```
RX MEDLINE=91172125; PubMed=2005867;
RA Irie K., Araki H., Oshima Y.;
RT "Mutations in a Saccharomyces cerevisiae host showing increased
RL holding stability of the heterologous plasmid pSR1.";
RN Mol. Gen. Genet. 225:257-265(1991).
RC SEQUENCE FROM N.A.
RP STRAIN=S288c / FY1678;
RA Ayadi A., Bordonne R., Camasses A., Madania A., Poch O.,
RL Tarassov I.A., Winsor B., Martin R.P.;
RC Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL PROTEIN INVOLVED IN PLASMID MAINTENANCE WITH
SMP2.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO S.POMBE SPAC4G8.12C.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; X58121; CA441123.1; -
DR EMBL; U55020; AAC49635.1; -
DR EMBL; 275057; CA99355.1; -
DR PIR; S13750; S13750.
DR SGD; S0005875; SMP3.
KW Transmembrane.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 122 123 MO -> IK (IN REF. 1).
FT CONFLICT 163 163 E -> G (IN REF. 1).
FT CONFLICT 169 169 S -> R (IN REF. 1).
FT CONFLICT 279 279 V -> L (IN REF. 1).
SQ SEQUENCE 516 AA; 59900 MW; 8D8404622CB69534 CRC64;

Alignment Scores:
Pred. No.: 4.9 Length: 516
Score: 82.50 Matches: 71
Percent Similarity: 34.31% Conservative: 46
Best Local Similarity: 20.82% Mismatches: 125
Query Match: 3.99% Indels: 99
DB: 1 Gaps: 15

US-09-814-661A-1 (1-1158) x SMP3_YEAST (1-516)
QY 369 GCATAACATATCGTTACTGTTTGGACATCCCGCTTTCGCCGATTCGCCCTCAGCG 428
Db 13 AlaIleGlyLeuMetLeuCysLeuGlyProSerTyrIleHisProAspGluHisPheGln 32
QY 429 GGTATAAAGAGATCTTTTTCCTGCGCTGCTCCATTTTAAATGCTTTATCT 488
Db 33 CysIleGlu-----IleLeuAlaMetGlnPheMetLysValLysGly 46
QY 489 GCT-----CCTTTGTGATCTTACGGT----- 509
Db 47 ThrIleProTrpGluPheLysSerLysPheAlaIleArgSerTyrGlyProLeuLeuLeu 66
QY 510 -----CTCACTAACCTCTCTCACTGCTCAATAATTTCCCGCTATGCAA 554
Db 67 ValTyrGlyProLeuPheThrIleLeuGluSer-----PheProGluIleGln 82
QY 555 AATTCCTCAAGACTACTTTTACCTCTCAAAATCCG----- 587
Db 83 AspAsnProAlaLeuIleLeuTyrSerMetArgLeuGlnAsnTyrValMetTyrIleLeu 102
```



```
QY 1008 GGAAGAGAAATGCACGGAGAGAAATGTTTCGAGAAATGACAACAATAGTAGCAGAGAGT 949
Db 85 ---LysAsnAsnThrIleProAlaAsnTrpLysGluInLysArgLysIleAspGluSer 103
QY 948 -----CCCTCAAAAGAGGTATCTAAGAGAAGAAAAGACAGAACAGTAGTGGGAAATGG 895
Db 104 IleArgArgLeuGluAlaValLeuLysGlnGluSerAsnArgIleGlnGluArg 123
QY 894 AAAGAGAAAAGAAAGAG----- 877
Db 124 LysGluLysGluGluGluAsnGlyProSerLysAlaLysSerAsnArgThrLysGluGln 143
QY 876 -----TATGAAAGGAACCTTTAGAAAGTCCATTTCCTCGACCTACCT 835
Db 144 GlyTyrPheGluGlyAsnAsnSerArgAsnIleProProProProProProProPro 163
QY 834 GG-----TTGAACATAGATTTTATGTCGCCA---GAACCAAAATTTGTTGTTGTCATG 785
Db 164 LysProProLeuAsn---AspProSerAsnProValSerLysAsnValAsnLeuPhe--- 181
QY 784 TCATGATCGATAGAGTTGAGTCTCTCCCTCCAAATCTTTTCCACATTTCTAATGAAGAA 725
Db 182 -----GlnIleGlyLeuThrPhePheLeuLeuSerPheLeuLeuAspLeuLeuAsnSer 199
QY 724 GCGGAGCAGAGAGCGAGCTGCCATGGAGTGTTCAGTAGACACATAGGAACCTCAGCC 665
Db 200 LeuGluGluGlnSerGluIleThrTrpGlnAspPheArgGluLysLeuLeuAlaLysGly 219
QY 664 ATAGGTGCAAGAGCAGCCCTTCTAAATTCGCCATGTGTCAGGTACGC-----AAT 614
Db 220 TyrValAlaLysLeuIleValAlaAsnLysSerMetValLysValMetLeuAsnAspAsn 239
QY 613 GTGGAAGGGCTGTGTGTGTGTCGAGCGATTTTCAGCGTAAAGTACTCTTGGGAATTT 554
Db 240 GlyLysAsnGlnAlaAspAsnTrpArg-----AsnPheTyrTyr 253
QY 553 TGCATAGCGGGAATATTATGACAGCTTGAAGAGAGGTTAGTACAGCCGTAAAGATCACAAA 494
Db 254 PheThrIleGlySerIleAspSerPheGluHisLysLeuGlnLysAlaGlnAspGluLeu 273
QY 493 GGAGCAGATAGACATTTAAANAATGGAGGAGGACAGCCAGGAAAAAAGATCTCTTTT 434
Db 274 AspIleAspLysAspPhe-ArgIlePro-----Valle 284
QY 433 ATACCGCTGAGCGGGAATCGGGGAAACGGCGATGTTCCAAACAGTACGATATGTT 374
Db 284 utyrValGlnGluGlyAsnTrpAlaLys---AlaMetPheGlnIleLeuProThrValle 303
QY 373 A-----TGGCTATTCTCTTTTGGGACGGATGACTTCCCAAAATGGC 332
Db 303 uMetIleAlaGlyIleIleTrpLeu-----ThrArgArgSerAl 316
QY 331 CACCGCGCGGGTAGT-----GGAGCCGCTTTGGGCTCAGCGGAGTAGTACAAAAA 278
Db 316 aGlnAlaAlaGlyGlySerArgGlyIlePheGlyLeuSerArgSerLysAlaLysLys 336
QY 277 ATTT 274
Db 336 aphe 337
```

RESULT 14

```
MIG1_KLUOLA
ID MIG1_KLUOLA STANDARD; PRT: 474 AA.
AC P50898;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Regulatory protein MIG1.
GN MIG1.
OS Kluyveromyces lactis (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
```

```
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;
RA MEDLINE=95402208; PubMed=7672126;
RA Casart J.-P., Georis I., Oestling J., Ronne H., Vandenhaute J.;
RT "The MIG1 repressor from Kluyveromyces fragilis: cloning, sequencing
RT and functional analysis in Saccharomyces cerevisiae.";
RL FEBS Lett. 371:191-194(1995).
CC -!- FUNCTION: INVOLVED IN GLUCOSE REPRESSION OF GLUCOSE METABOLISM
CC GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE CREB/MIG GROUP OF C2H2-TYPE ZINC-
CC FINGERS PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z50017; CAA90320.1; -
CC HSSP; P08047; 1SP2.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2.2
CC PRINTS; P000048; ZINC_FINGER..
CC ProDom; PD000003; Znf_C2H2.2.
CC SMART; SM00355; Znf_C2H2.2.
CC PROSITE; PS00028; ZINC_FINGER_C2H2.2.
CC PROSITE; PS0157; ZINC_FINGER_C2H2.2.
CC Transcription regulation: Repressor; DNA-binding; Nuclear protein;
CC Zinc-finger; Metal-binding; Repeat; Carbohydrate metabolism.
CC ZN_FING 26 48
CC ZN_FING 54 78
CC ZN_FING 217 225
CC DOMAIN 217 225
CC POLY-SER.
CC SEQUENCE 474 AA; 52907 MW; 4A5F3EA3603B17C3 CRC64;
CC
CC Alignment Scores:
CC Pred. No.: 6.83 Length: 474
CC Score: 81.00 Matches: 70
CC Percent Similarity: 34.29% Conservative: 38
CC Best Local Similarity: 22.22% Mismatches: 112
CC Query Match: 3.91% Indels: 95
CC DB: 1 Gaps: 11
CC
CC US-09-814-661a-1 (1-1158) x MIG1_KLUOLA (1-474)
```

```
QY 299 GAGCCCAACGGCTCCACTACCCCGCGGTTCGCCATTTTGGGAAGTCAVCCGTC--- 355
Db 151 GluProLeuLysSerLeuArgSerLysProMetPheAspLeuGlySerAspGluSerAsp 170
QY 356 -----CAAAAAGGAATAGCCATTAACATATCG 382
Db 171 GluCysGlyIleTyrSerValProProIleArgSerGlnAsnAsnSerGlyAsnIleAsp 190
QY 383 TTACTGTTTTGGAACATCGCCGTTTCGCCGATTCGCCCTCAGCGGGTATAAAGAGA 442
Db 191 LeuLeuLeu---AsnAlaAlaLysPheGluSerAspTysAlaSerSer----- 205
QY 443 TCTTTTTTTTCCCTGGCTGCTCCATTTTAAATCTTATCTATCTCTCTCTCTCTCTCTCT 502
Db 206 SerPheLysPheIleAsp-LysLeuProLeuThrSerSerSerSerSerProSerLeuSe 225
QY 503 TTAC----- 506
Db 225 rPheThrSerHisSerIleAsnAsnSerSerSerGlyLeuLeuLeuProArgProAlaSe 245
QY 507 -----GCTCTCACTACCTCTCTTCAACGCTCAATTAATTCCTCGCTATGCAAAATTC 559
Db 245 rArgAlaLysLeuSerAlaLeuSerLeuGlnArgMetThrProLeuSerGlnAsnSe 265
```

THIS PAGE BLANK (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 11:53:09 ; Search time 51 Seconds
(without alignments)
4365.634 Million cell updates/sec

Title: US-09-814-661A-1
Perfect score: 2069
Sequence: 1 aatgagcaacccgtgtcaaca.....cctcttggatcaatgctt 1158

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xip
-Q=Cgn2.1/USPTO.spool/US09814661/runat_16072003.115259.14538/app.query.fasta_1.1351
-DB=PIR_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661 @CGN_1_1_64 @runat_16072003.115259.14538 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	537	26.0	104	2 S49803	hypothetical prote
2	220	10.6	68	2 S78729	protein YML058w-a
3	102.5	5.0	835	2 T15177	hypothetical prote
4	101	4.9	307	2 A71605	probable integral
5	99.5	4.8	347	2 F22845	hypothetical prote
6	96	4.6	493	2 T34453	hypothetical prote
7	95.5	4.6	164	2 S19370	probable membrane
8	94.5	4.6	885	2 S61189	probable membrane
9	92.5	4.5	508	2 E90594	amino acid permeas
10	92.5	4.5	654	2 S69673	SAC7 protein - yea
11	91	4.4	104	2 S64045	probable membrane
12	90.5	4.4	531	2 T11074	NADH2 dehydrogenas
13	90	4.3	501	2 T23011	hypothetical prote
14	90	4.3	902	2 S49931	SyG1 protein - yea

15	89	4.3	288	2 A05235	hypothetical prote
16	89	4.3	509	2 F71526	hypothetical prote
17	89	4.3	762	2 G88436	protein T04A8.13 l
18	88.5	4.3	2212	2 T28157	erythrocyte membra
19	88	4.3	112	2 S48472	probable membrane
20	87.5	4.2	105	2 B30010	hypothetical ORF-5
21	87.5	4.2	410	2 T11064	NADH2 dehydrogenas
22	87	4.2	533	2 S62489	hypothetical prote
23	86.5	4.2	354	2 B88650	protein C09G12.4 l
24	86	4.2	374	2 T25943	hypothetical prote
25	86	4.2	1786	1 H71527	probable excinucl
26	85.5	4.1	311	2 T31976	hypothetical prote
27	85.5	4.1	2437	2 T18482	hypothetical prote
28	85	4.1	214	2 A35349	H+-transporting AT
29	85	4.1	372	2 AE1370	S. pyogenes Rofa r
30	85	4.1	791	2 T24435	hypothetical prote
31	85	4.2	1711	1 A47392	chromodomain-helic
32	85	4.2	3255	2 G81702	adherence factor T
33	84.5	4.1	285	1 C70066	conserved hypoteth
34	84.5	4.1	419	2 E81327	probable integral
35	84	4.1	1983	2 T00385	KIAA0624 protein -
36	83.5	4.0	504	2 AC1740	S. pyogenes Rofa r
37	83	4.0	153	2 T12419	NADH2 dehydrogenas
38	83	4.0	343	2 T22966	hypothetical prote
39	83	4.1	672	2 T21469	hypothetical prote
40	83	4.1	1431	2 A97196	hypothetical prote
41	82.5	4.0	337	2 T23892	hypothetical prote
42	82.5	4.0	344	2 T33909	hypothetical prote
43	82.5	4.0	516	2 S67037	SMP3 protein - yea
44	82.5	4.0	825	2 S54465	YTA12 protein prec
45	82	4.0	320	2 S18444	G protein-coupled

ALIGNMENTS

RESULT 1

S49803
hypothetical protein YML058w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMG958.04
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 19-Apr-2002
C:Accession: S49803
R:Devlin, K.; Churcher, C.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49800
A:Accession: S49803
A:Molecule type: DNA
A:Residues: 1-104 <DEV>
A:Cross-references: EMBL:246729; NID:9577134; PIDN:CAA86717.1; PID:9577138; GSPDB:GNO
C:Genetics:
A:Gene: SGD:SML1; MIPS:YML058w
A:Cross-references: SGD:S0004523
A:Map position: 13L
C:Superfamily: Saccharomyces hypothetical protein YML058w

Alignment Scores:
Pred. No.: 8.81e-47 Length: 104
Score: 537.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.95% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x S49803 (1-104)

Qy	549	ATGCAAAATTCCTCCAGACTACTTTACGCTCAAAATCGCTGCCAACACACACAGCCCTC	608
Db	1	MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnAlaPro	20
Qy	609	TCCACATTGCGTACCGTGACCATCGCGAATTTAGAGGGTGCCTTTGCCACCTATGCT	668
Db	21	SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla	40

QY 669 GAGTTCCTATGTTCTACTCAAACTCCAGGCGCTCGCTTCTCGCTCGCTTCT 728
Db 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSer 60
QY 729 TCATTAGAAATGGGAAAGGATTTGGAGGAGACTCAACTCTATCGCATCATG 788
Db 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspMet 80
QY 789 ACAACAACAATTTGGTCTGCGGAACATAATCTATCTTCAACGAGGTAAAGTCGAG 848
Db 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
QY 849 GAAATGGACTTC 860
Db 101 GluMetAspPhe 104

RESULT 2

S78729
protein YML058w-a - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 15-Jan-1999 #sequence_revision 15-Jan-1999 #text_change 15-Jan-1999
C:Accession: S78729
R:Devlin, K.; Churcher, C.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49800
A:Accession: S78729
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-68 <DEV>
A:Cross-references: EMBL:Z46729; MIPS:YML058w-a
C:Genetics:
A:Map position: 13L

Alignment Scores:
Pred. No.: 3.64e-14 Length: 68
Score: 220.00 Matches: 43
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.63% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x S78729 (1-68)

QY 2 ATGAGCAACCGTGTCAACAGAGTGTCAAGACGGCTACTATTCCCGAAGGATCACGTT 61
Db 26 MetSerAsnArgValAsnLysSerValLysThrGlyTyrLeuPheProLysAspHisVal 45
QY 62 CTTTCTGCCAACATCATGCGTCGGAACGTGCGGGCGGTCTTTCTGACATTGCTAAGAAT 121
Db 46 ProSerAlaAsnIleIleAlaValGluArgGlyGlyLeuSerAspIleGlyLysAsn 65
QY 122 ACTTCCAAC 130
Db 66 ThrSerAsn 68

RESULT 3

T15177
hypothetical protein C48E7.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15177
R:Wamsley, P.; Kramer, J.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C48E7.
A:Reference number: Z18303
A:Accession: T15177
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-835 <WAM>
A:Cross-references: EMBL:AF000262; NID:g1947126; PID:g1947134; PIDN:AAB52939.1; GSPDB:CN
A:Experimental source: strain Bristol N2, clone C48E7
C:Genetics:
A:Gene: CESP:C48E7.8

A:Map position: 1
A:Introns: 56/1; 150/1; 197/3; 240/1; 406/2; 456/1; 497/3; 535/2; 679/2; 734/1
Alignment Scores:
Pred. No.: 0.0739 Length: 835
Score: 102.50 Matches: 81
Percent Similarity: 38.35% Conservative: 49
Best Local Similarity: 23.89% Mismatches: 109
Query Match: 4.95% Indels: 100
DB: 2 Gaps: 17
US-09-814-661A-1 (1-1158) x T15177 (1-835)
QY 212 ATGTTTATGTTAGACCTCGTACATAGTAATAATTGACGTTT---TTTGGCCAAC 268
Db 495 LeuTyrLeuThrLeuProIleVal-IlePhePheLeuSerLysPheArgPheGlyPheTh 514
QY 269 ATTTGAAATTTTTTTTGTACCTCGCGTGGCCCAACACGGGCTCCATACCCG----- 323
Db 514 rPheAlaIleValLeuCysIleLeuAlaLeuValTyrArgSerLeuMetTyrAlaIleTy 534
QY 324 -----CCGCGTCCCATTTTGGGAACATCTCCGTCCTCCCAAAAGGAATAGCC---AT 373
Db 534 rThrLeuProAlaAsnMetPheValGluMetLeuSerGlyLysPheGluLysAlaGluLe 554
QY 374 AACATATCGTTACTGTTTGGGAACATCGCGTTCGCCCGCATTCGCCCTCAGCGGGTAT 433
Db 554 userPheLysLeuLeuTyr---ThrSerProLeuSer-ArgCysProPro----- 569
QY 434 AAAAGAGATCTTTTTTTTTCCTGGCTGTCCCTTCCATTTTAAATGTCTTATCTCTCC 493
Db 570 -----PhePheLeuGlyAsnLeuPhe-----AsnTyrIleThrL 581
QY 494 TTTGTGATCTTACGGTCTCAGCTTCACTCTTCAACTGCTCAATAATATTCGCGCTATGCA 553
Db 581 euMetProArgAlaIle-----LeuGlyIleLeuThrGlyTrpTyrIleCysG 597
QY 554 AAATT-----CCCAAGACTACTTTTACGCTCAAAATCGCTGCCA 592
Db 597 inIleArgTyrLeuGlyThrLeuArgSerIleThrSerPheGlyLeuLysIle-LeuAla 616
QY 593 ACAACAACAGCCCTTCCACATTCGGTACCGTGACCATGGCGAATTTAGAAGGTGCC 652
Db 617 ValSerLeu-----LeuAlaPheAlaLeuPheGlyProPhePheTyrSerGlu 632
QY 653 TTTGCCACCTATGCGTGAGGTTCCTATGTTGCTACTCAAACTCCATCGGACGCTCCCG 712
Db 633 PheLeuThrTyr-----ProHisAlaIleSerAsnArgIleLeuTrpAlaSerGly 649
QY 713 TTTCTGCTCCGCTTCTTCATTAGAAATGTGGGA-----AAAGGATTT 754
Db 650 LeuAlaIleLeuValValLeuAlaGlnAspGlyTyrAsnPheGlyValPheLysValPhe 669
QY 755 GGAGGAGAGACTCAACTCTATCGATCATGACATGAACAACAACAAATTTGGTTCGGCGA 814
Db 670 GlyGlyGlnThrLeuVal----- 675
QY 815 ACTAAATCTATGTTCAACACCGGTAAGGTTCAGGAAATGGAGTCTTAAGTTCTTTTCA 874
Db 676 -----ValLeuSerArgLeuSer 691
QY 875 TACTCTTTTCTTTCTTTTCCATTTCCCACTAGTTCCTCTCTTT---TCTTCTCTAGA 931
Db 682 PheGlyValTyrLeuSer---HisGluProIleLeuLeuTyrTyrLeuAsnSerLeuArg 700
QY 932 TACCCTTCTTTTTCAGGAGACTCTCGTCTCTACTATTGTTGTCTCTCGAACAATCTCTCC 991
Db 701 GlnProMetSerProThrSerPheGlyTyr----- 710
QY 992 CGTGCATTTTCTTTTCCCTTTTATATACATATATATATATATATATATATATATCTCT 1051
Db 711 -----PheMetPheIleThrIleSerIleTyrValLeuSerLeuValCysAla 726

1052	QY	ACGATATTTTGTATTTCTGTCGCTTTA-----TCAAGATAGATGCTAT	1093
727	Db	PhePheIleAlaIleSerIleGluValProLeuLeuAsnLeuGluArglyLeuMet	746
1094	QY	ATACG-----TTTGATACAGCTAGATATCGTAGCGCAACAT	1132
747	Db	ThrTrgArgasnMetLysGluAspThrMetLysAspArgMetGluArgIhis	763

RESULT 4

probable integral membrane protein PFB0845w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: A71605
 R:Caradner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1128-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: A71605
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-307 <CAR>
 A:Cross-references: GB:AE001362; NID:g3845287; PIDN:AA71958.1; PID:g3845287
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0845w

Alignment Scores:	
Pred. No.:	0.0873
Score:	101.00
Percent Similarity:	50.57%
Best Local Similarity:	33.33%
Query Match:	4.88%
DB:	2
Length:	307
Matches:	29
Conservative:	15
Mismatches:	21
Indels:	22
Gaps:	4

US-09-814-661A-1 (1-1158) x A71605 (1-307)

[illegible]

RESULT 5

F22845
hypothetical protein 6 - Trypanosoma brucei mitochondrion
C:Species: mitochondrion Trypanosoma brucei
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-May-2000
C:Accession: F22845
R:Hengstens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.
Nucleic Acids Res. 12; 7327-7344, 1984
A:Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift con-
ondrial maxi-circle DNA.
A:Reference number: A93537; MUID:85037915; PMID:6093040
A:Accession: F22845
A:Molecule type: DNA
A:Residues: 1-347 <H>
C:Genetics:

A;Genome: mitochondrion
A;Genetic code: SGC6
C;Superfamily: cytochrome
C;Keywords: mitochondrion

Alignment Scores:	
Pred. No.:	0.128
Score:	99.50
Percent Similarity:	45.74%
Best Local Similarity:	32.98%
Query Match:	4.81%
DB:	2
Length:	347
Matches:	31
Conservative:	128
Mismatches:	32
Indels:	19
Gaps:	5

US-09-814-661A-1 (1-1158) x F22845 (1-347)

853 TGGACTTCTAAGTTCCATTCATACTCCTTTCTCTTTCTCATATTCCCACTAGTTCT 912 QY
||| ::: ||||||||::: |||:|||| |
22 TrpTrpPheAspLeuAspPhelIeLeuTyraSPheValPheAspPhelValCysIle 41 Db
||| ::: ||||||||::: |||:|||| |
913 GTTCCTTTCTCTCTTAAGATA-----CCCTCTTTTTGAGGAGCTCTC 954 QY
||| ::: |||

42	ThrPheIleValPheIleValLeuGlyPhePheIleAlaGlyIlePhePheSerPheValPhe	61
955	GTCTACTATTGTGCA---TTCTCGAAACATCTCTCCCGTCAGTCTTCTTCCCTT	1011
QY		
62	ValLeuLeuPheIleThrPheGlyIleCysSerLeuThrMetLeuPheThrGlyTyr	81
QY	TATATACATATATATATATA-----TATATATGCTCTCTACGTATTT---	1059
82	TyrIleTyr---TyrIleTyrIleLeuTyrAsnPheIleCysPhePheAlaPheGly	100
QY	-----TTGATATTTCTGTGCTTTATC	1080
101	IleAsnPheLeuIleTyrTyrIleGluPheIlePheIle	114
Db		

RESULT 6

hypothetical protein T19U12.8 - Caenorhabditis elegans

C; Species: *Caenorhabditis elegans*
C; Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C; Accession: T34453

submitted to the EMBL Data Library, April 1997

A: description: The sequence of C. elegans cosmid I19H12.
A: Reference number: 221528

A/ACCESSION: T34433
A/STATUS: preliminary; translated from GB/EMBL/DDAT

A;Molecule type: DNA
A:Residues: 1-103

A; Cross-references: EMB
A; Experimental course:

C;Genetics:

A;Map position: 5

Pred. No.:	0.313	Length:	493
Score:	95.00	Matches:	43
Percent Similarity:	43.6%	Conservative:	19
Best Local Similarity:	30.2%	Mismatches:	58
Query Match:	4.64%	Indels:	22
DB:	2	Gaps:	6

US-09-814-661A-1 (1-1158) x T34453 (1-493)

Qy 360 AAGAAATAGCCATACATATCGTTACTGTTTGGAAACATCGCCCGTTTCGCCGATTCC 419

[illegible][illegible]


```

Db 433 LeuIleLeuPheGluSerLeuAsnSerThrValLeuIleProLeuArgAspGluHis 452
QY 515 TAACCTCTCTCAACTG---CTCAATAATTCCCGCTATGCAAAATCCCAAGACTACT 570
Db 453 LeuGlnLeuPheAsnValLeuPheAsnTyrLeuProLeuLeuLysSerAsnLeuThr 472
QY 571 TTTAGCCTCAAAATCGCTGCC----- 591
Db 473 LeuArgPheGluIleGluLysArgSerMetCysAsnIleGluThrLeuArgPheIleSer 492
QY 591 ----- 591
Db 493 LeuTrpLysTyrTyrGlnPheValMetAspThrSerSerSerSerPheValIleAsp 512
QY 592 -----AACAAACAAGCCCTT 609
Db 513 TyrAspGluAspMetHisMetAlaCysLeuLeuSerLeuAsnHisGluThrGlnAsnGln 532
QY 610 CCACATTCGGTACCGTACCATGGCGGAATTA-----GAAGGTGGCTTTGGCCACTA 663
Db 533 SerHisIleLeu-----ThrTrpAsnPheIlePheLysAsnTyrCysTrpArgHisLeu 550
QY 664 TGGCTGAGGTTCTCTACTACTCAAACTCCATCG-----GCAGCTCCGCTTCTG 717
Db 551 PheLeuGlnLeuProLeuLeuMetSerGluProPheThrAsnSerThrProIleIle 570
QY 718 CTCCGCTTCTTCAT-----TAGAAATGTGGGAAAGG 750
Db 571 AspProLeuLeuAsnAsnAspPheGluLeuIleAspPheGluValAsnLeuMetLysTyr 590
QY 751 ATTTGGAGAGAGACTCACTCATGATCATGACATGAACAACAACAAATTTGGTTCTG 810
Db 591 LeuGlnSerLysAspGlnLeuSerIleAspLysIleIleGlnLeuIleLysLeu 610
QY 811 CGCAACTAAATCTA-----TGTTCAACAGGTAAGTCGAGGAAATGG 855
Db 611 LysAsnLysAsnIleGluValSerGlnGlyCysLeuThrThrProSerIleIle-AsnAs 630
QY 856 ACTTCTAAAGTTCCCTTCATAC-----TCCTTTCTTTCTCTTCCATTTCCCACTAGT 909
Db 630 nIleMetAspSerLeuIleTyrArgAsnSerMetLeuTyr---LeuAsnPheTyrLeuLe 649
QY 910 TCTGTTCTTTCTCTCTTAGA---TACCTTCTTTTTCAGGAC----- 950
Db 649 uLeuGlnPheGluThrLeuLysAsnTyrAlaLys-PheAsnGluIleLeuGluAspPheL 669
QY 951 -----TCTCGTCTACTATTGTTGTCATCTCGAACATCTCTCCGCTGCATTTTC 1002
Db 669 euGluSerArgGluThrLeuPhePheValPheSerAsnLeuAlaAsnIleLysPheA 689
QY 1003 CTTTCCCTTTATATATATATATATATATATATATATATATATATATATATATAT 1062
Db 689 laGlyHisGluPheThrPheIleAsnLysSerIleValValLeuGlnThrLeuValLeu 709
QY 1063 TATTTCTGTCTTTTCAAGA 1085
Db 709 etLeuLeuAlaLeuTyrGlnArg 716

```

RESULT 9

E90594
amino acid permease MYPU_6610 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: E90594
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90594
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <KUR>

A:Cross-references: GB:AL445566; PID:g14090076; PIDN:CAC13834.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU_6610
A:Genetic code: SGC3
Alignment Scores:
Pred. No.: 0.721 Length: 508
Score: 92.50 Matches: 89
Percent Similarity: 33.77% Conservative: 66
Best Local Similarity: 19.39% Mismatches: 133
Query Match: 4.47% Indels: 171
DB: 2 Gaps: 20

US-09-814-661A-1 (1-1158) x E90594 (1-508)

```

QY 60 TTCCTTCTGCCAACATCATTTG-----CCGTCGACGTCGCGGG 98
Db 97 PhePheIleProValSerLeuPheThrMetThrIleTyrValThrSerSerPheValAsp 116
QY 99 GTCTTCTCTGACATTG-----GTAACAATACTTCCAACCTAAGAGCATGCTTCTC 146
Db 117 ValGlyValSerLeuLysAsnGlyTrpLeuValLeuPheAlaPheLeuIlePheLeu 136
QY 147 TTTTCTTTTGTAGGCCAATGATFAGGAAAGAACAAATAGATTATATACTCAGAAATATAG 206
Db 137 PhePheMetVal-----IleAsnLeuIleSerIleLysIleGly 149
QY 207 TAGATATGTTTATGTTTGTAGACCTCGTACATAGGAATAATGACGTTTGTGGCCCA 266
Db 150 GluValAlaGlnTrpIle---ThrThrIleValLysValValProLeuPheValLeuPro 168
QY 267 ACATTTGAAATTTTGTGTTTACCTCGCGCTGAGCCCAACGCGCTCCACTACCGCGCG 326
Db 169 ValIleAlaPheValPheAlaAspLeuGluLeuGlyAsnThrPheLeuGln----- 185
QY 327 CGGTCGCCATTTGGGAAGTCACTCGTCCCAAAAAGAAATAGCCATACATATCGTTAC 386
Db 186 -----LysGlnIleLysPro-----GluValGlyIleThr----- 195
QY 387 TGTTTTGGAACTCCCGCTTTTCGCGCGATTCGCGCTCAGCGGGTATATAAAGAGATCTT 446
Db 196 -----GlyMetSer----- 198
QY 447 TTTTCTCTGCTGCTCCCTTCCATTTTAAATGTCTTATCTGCTCTCTTGTGATCTTAC 506
Db 199 -----LysTrpLeu-----IleIleIleAlaGlyLeu----- 207
QY 507 GGTCTCACTAACTCTCTTCAACTGCTCAATTTCCGCTTATGCAAAATTTCCCAAGAC 566
Db 208 -----ProAlaIleThrPheAlaTyrAsp 215
QY 567 TACTTTTACGCTCAAAATCGCTGCCAACAAACAAAGCCCTTCCACATGCGTACCCTG 626
Db 216 AsnPheTyrAlaIleSerAsnIleLysGluGluLeuSerProLysAlaGluLysIle 235
QY 627 ACCATGGCGGAATTTAGAAGGCTGCTTTGCCACCTATGCTGAGTCTCTCTATGTGTCT 686
Db 236 GlyMetAlaIleVal-----IleGlyLeuAlaIleIleThrValIleTyrMetLeuIle 253
QY 687 ACTCAAAACTCCATGGCAGCTCCGCTTCTGCTCTCCGCTTCTTCAATAGAAATGTGGAA 746
Db 254 ThrValAlaPheAsnTyrGlySerAsnAspGlyThrIleAsnThrIleSerPheLeuGlu 273
QY 747 AAGGATTTGGAGGACAGACTCAACTCTATCGATCATGAC----- 785
Db 274 LysProGluAsnAlaArgLeuPheAlaPheAsnSerCysIleAlaIleGlyIleLeu 293
QY 786 -----ATGAACAACAACAATAATTTGGTTCTGCG---GAACATAAATCTATGTTCAACCCAG 836
Db 294 GlyIleIleAsnSerLeuAlaLeuSerSerProTyrGlnLeuArgGlyLeuTyrGluGln 313
QY 837 GGTAAAGTCGAGGAAATGGACTTCTTAAAGTTCCTTT----- 872

```


F:23-39/Domain: transmembrane #status predicted <TM1>
F:72-88/Domain: transmembrane #status predicted <TM2>

Alignment Scores:
Pred. No.: 0.763 Length: 104
Score: 91.00 Matches: 26
Percent Similarity: 46.74% Conservative: 17
Best Local Similarity: 28.26% Mismatches: 17
Query Match: 4.40% Indels: 32
DB: 2 Gaps: 4

US-09-814-661A-1 (1-1158) x S64045 (1-104)

```
QY 871 TTCATACCTCTTTCTTTCTTTCCATTTCCCACTAGTTCGTCTCTTTCTCTCTAG 930
      |||||:||||| ||| |||||
Db 30 PheIleValPheSerLeuAlaGluValProSerArgLeuThrAsnPhePheSer--- 48
      |||||:||||| ||| |||||
QY 931 ATACCTCTCTTTTCAGGGACTCTCGCCCTACTATTGTTGTCATTCTCGAAACATCT--- 987
      |||||:||||| ||| |||||
Db 49 -----IleMetIleLeuLeuThrPheSerAsnPheSerGln 60
      |||||:||||| ||| |||||
QY 988 -----CTCCGTCGCAATTTCCCTTTTATATACATATATATATATATATATATA 1041
      |||||:||||| ||| |||||
Db 61 AsnIleArgProArgIleTyrLeu-----IleHisGluPheLeuHisIleTyrVal 77
      |||||:||||| ||| |||||
QY 1042 TGCTCTCTCTACGTATTTTGTATTCTGTCTTTATCAAGATAGTCTATAATACGTT 1101
      |||||:||||| ||| |||||
Db 78 CysIleTyrPheValIleArg----- 84
      |||||:||||| ||| |||||
QY 1102 TGATACAGTAGATATCGTAGCGCCCAACATTTGCTCC 1137
      |||||:||||| ||| |||||
Db 85 -----LeuSerValValProArgLeuSer 92
      |||||:||||| ||| |||||
```

RESULT 12

TM1074
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - nematode (Onchocerca volvulus) m
C:Species: mitochondrion Onchocerca volvulus
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
C:Accession: T11074
Mol. Blochem. Parasitol. 95, 111-127, 1998
R:Keddie, E.M.; Higazi, T.; Unnasch, T.R.
A:Title: The mitochondrial genome of Onchocerca volvulus: Sequence, structure and phylog
A:Reference number: 217246; MUID:98434257; PMID:9763293
A:Accession: T11074
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-531 <RED>
A:Cross-references: EMBL:AF015193; NID:g2735934; PID:g2735946; PIDN:AAC61621.1
A:Experimental source: strain forest
C:Genetics:

A:Genome: mitochondrion
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
Pred. No.: 1.17 Length: 531
Score: 90.50 Matches: 69
Percent Similarity: 35.89% Conservative: 34
Best Local Similarity: 24.04% Mismatches: 83
Query Match: 4.37% Indels: 101
DB: 2 Gaps: 14

US-09-814-661A-1 (1-1158) x T11074 (1-531)

```
QY 273 GAAATTTTGTGTTTACCTCGCGTCGAGCCCAACGGCTCCACTACCGCGCGGTGCG 332
      |||||:||||| ||| |||||
Db 263 GlnIleGlyPheCysPheLeuAlaIleGly---SerGlyLeuHisTyr---LeuSerTyrV 281
      |||||:||||| ||| |||||
QY 333 CCATTTTGGGAAGTCATCGTCCCAAAAAGAAATAGCCATAACATATGCTTACTGTTT 392
      |||||:||||| ||| |||||
Db 281 alHisMetIleSerHisSerPhePheLys-----SerLeuLeuPheM 295
      |||||:||||| ||| |||||
QY 393 GGACATCGCCCGT-----TTCGCCCGATTCGCGCTCAGCGGGTATAAAAGAGATCTT 446
      |||||:||||| ||| |||||
```

```
Db 295 etGlnMetGlyTyrLeuIlePheIleAsnPheGlyGlnGlnAspTyrArgGlyTyrSerP 315
      |||||:||||| ||| |||||
QY 447 TTTTTCCTGCTGCTCCCT-----TCCATTTTAAATGCTTATCTGCT 491
      |||||:||||| ||| |||||
Db 315 hePheGlyPheCysAlaProValLeuValGlnLeuGlnIlePheLeuSerValPheCysL 335
      |||||:||||| ||| |||||
QY 492 CCTTTGTGATCTTACGGTCTCACTAACCTCTCTTCAACTGCTCA-----ATAATT 542
      |||||:||||| ||| |||||
Db 335 euCysGlyLeuLeuPheThrSer-----GlySerCysSerLysGluTyrPheMetS 352
      |||||:||||| ||| |||||
QY 543 CCGGCTATGCAAAATTCACCAAGACTACTTTTACGGCTCAAAATCGCTGCCAACACAA 602
      |||||:||||| ||| |||||
Db 352 erArgPhe----- 354
      |||||:||||| ||| |||||
QY 603 GCCCCTTCACATTCGGTACCGTGACCATGGCGGAATTTAGAAGGTGCTTTGCCACCT 662
      |||||:||||| ||| |||||
Db 355 -----TyrTyrAspSerTyr----- 359
      |||||:||||| ||| |||||
QY 663 ATGGCTGAGGTTCTCTATGTTGTCTACTCAAACTCCATGGGCAG---CTCCGCTTCTGCC 719
      |||||:||||| ||| |||||
Db 360 -----GlyPhePheLeuValPhePheTyrPhePheGlyValPheLeuThrPheCys 377
      |||||:||||| ||| |||||
QY 720 TC-----CGCTTCTTCATTAGAAATGTGGAAAAGGATTTGGAGGAGAGACTCAAC 770
      |||||:||||| ||| |||||
Db 377 yrCysTyrArgMetPhePheLeuPheArgValGly---AlaPheGly----- 391
      |||||:||||| ||| |||||
QY 771 TCTATGCATCATGACATGAACAACAATAATTTGGTTGCTGGCGAATAAAATCTATGTT 830
      |||||:||||| ||| |||||
Db 391 ----- 391
      |||||:||||| ||| |||||
QY 831 AACAGGGTAAGGTCGAGGAAATGGACTTCTAAAGTTCTCTTATCTACTCTTTTCTTTCT 890
      |||||:||||| ||| |||||
Db 392 -----PheAspTyrValGlyPheSer---SerLysLe 401
      |||||:||||| ||| |||||
QY 891 CTTTCCATTTCCCACTAGTTCGTCTTCTTCTCTTAGATACCTCTCTTTTTCAGGGAC 950
      |||||:||||| ||| |||||
Db 401 upHeTyrPheSerCysPhePheLeuValPhePheSerValVal-----PheThrPheTr 419
      |||||:||||| ||| |||||
QY 951 TCTCGCTACTATTCTGTGTCATTCTCGAAACATTTCTCCCGTGCATTTTCTCTTCCCT 1010
      |||||:||||| ||| |||||
Db 419 prPValPheSerLeuLeuSerPheSerValAlaPheAsnArgPheGluPheLeuValVa 439
      |||||:||||| ||| |||||
QY 1011 TTATATACATATATATATATATATATATATATATATATATATATATATATATCTG 1070
      |||||:||||| ||| |||||
Db 439 lTyrPheTyrLeuPhe---PheValTyrCysPheCysGlyTyrPhePheArgTyrPheVa 458
      |||||:||||| ||| |||||
QY 1071 TGCTTTTATCAAA 1083
      |||||:||||| ||| |||||
Db 458 lValGluPheLys 462
      |||||:||||| ||| |||||
```

RESULT 13

T23011
hypothetical protein F59F4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T23011
R:Harris, B.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19652
A:Accession: T23011
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-501 <NIL>
A:Cross-references: EMBL:Z81095; PIDN:CAB03159.1; GSPDB:GN00028; CESP:F59F4.3
A:Experimental source: clone F59F4
C:Genetics:
A:Gene: CESP:F59F4.3
A:Map position: X
A:Introns: 26/2; 62/2; 87/3; 119/1; 135/2; 172/1; 220/1; 243/2; 306/1; 365/2; 422/2;
C:Superfamily: Caenorhabditis elegans hypothetical protein F59F4.3

Alignment Scores:

Db		94	eValIleTyrPheLeuPheCys----	Phe-IlePhePheValIleArgCysLeuPheIleV	113
QY		501	TCTTACGGTCCTCAATAACCTCTCTTCAAC-----	TGCT	533
			:::::		
Db		113	alleMetAspPhe--LeuPhePheAsnPheAsPIlePheValSerIleLeuLeuCysA	132	
QY		534	CATATAATTTCCCGGTATGC AAAATTCCGAAGACTACTTTTTACGCTCAAATCGCTGCCAA	593	
			:		
Db		132	spileVal-----TyrLeuAspPheIleSerLeuLeuLeuLeuTyr-----	145	
QY		594	CAACAACAGCCCCCTTCCACATTTCGTCACCTGACCATTGGCGGAATTTAGAAGGTGCCT	653	
			:		
Db		146	-----PheAsnPheIlePheAsn-----	151	
QY		654	TTGCCACTATGGCTGAGGTTCTCATGTGTGTCTACTAAAACCTCCATGGCAGCTCCGCT	713	
				:	
Db		152	--PheIleTyrGlyPhePheSerPheValIleLeu-----	GlyLeuLeuLeuP	167
QY		714	TCTGCCCTCCGCTCTCTCATTAAGAATATGGGCAAAAGGATT-----	GGAG	758
			:: : :		
Db		167	heLeuLeuLeuPheLeuValIleAsnLeuPheGlyPheThrPheLeuValTyrGlyI	187	
QY		759	GAGAGACTCAACTCTATCGATCATCATGACAACAACAATTTGGTTCTGGCGAACTA	818	
			:::	:	
Db		187	leGlnIleLeuLeu-----TyrTyrValTyrTrpLeuTyrM	200	
QY		819	AAATCTATGTTCAACACCGGTAAAGTTCGAGGAATGGACTCTCA-----	862	
					:
Db		200	etIleTyr-----SerArgSerCysTyrIleLeuMetProAlaIleLeuI	215	
QY		863	-----AAGTTCTTTCATACACTCTTTCTCTTCTTCCATTCCCACCTAGTCTCTG	914	
			: : ::		:
Db		215	lepPheLysPheIleTyrPheAspValPhePheValPheValPheIleLeuLeuP	235	
QY		915	TC-----TTTTCTCTCTTAGATACCTCTCTTTTCAGGGACTTCGTCCTACTATT	965	
Db		235	heIleIleSerPhePheSerPhePheLeuLysAspPheLeuPheLeuSerLeuTyrPhe-	254	
QY		966	GTTGTCTATTCTCGAAAACTCTCTCCCGTGCAATTTCTCCCTTTATATACATATA	1021	
			:::	:	
Db		255	-----AspIlePheGlySerLeuTyrAsnTyrAsp-----	IleLeuS	267
QY		1026	TATATATATATATATATGTCCTCTACGTATTTTGTATTTCTGTGCTTTATCA	1081	
				:::	:
Db		267	erTyrSerIlePheTyrTyrGlnAsnAsnGlnPheCysLeuThrGlnLeuLeuSer	285	

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 14:31:22 ; Search time 3055 seconds
(without alignments)
11031.440 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 1158

Sequence: 1 aatgagcaaccgtgtaaca.....cctctcttgatcaatgcttt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1038	89.6	20951	8	SC9958	246729 S.cerevisia
2	691	59.7	2998	8	SCCMP2	X54964 Yeast CMP2
3	74	6.4	1767	6	I90044	I90044 Sequence 20
4	74	6.4	2353	6	I90040	I90040 Sequence 11
5	74	6.4	2353	8	YSCALA2	M64840 S.cerevisia
6	37	3.2	2134	10	D38434	D38434 Mus musculus
7	37	3.2	2648	10	AB001873	AB001873 Mus muscu
C 8	36	3.1	85338	3	AC004279	AC004279 Drosophil
C 9	36	3.1	122049	3	HSB166C10	AL121870 Human DNA
C 10	36	3.1	173380	3	AC099011	AC099011 Drosophil
C 11	36	3.1	181728	3	AC008003	AC008003 Drosophil
C 12	36	3.1	185469	2	AC008327	AC008327 Drosophil
C 13	36	3.1	307748	3	AE003578	AE003578 Drosophil
C 14	35	3.0	5348	3	AE003326	AE003326 Drosophil
C 15	35	3.0	44855	3	DMC52C10	AL035311 Drosophil
C 16	35	3.0	55452	3	AC004328	AC004328 Drosophil
C 17	35	3.0	156776	9	AC011443	AC011443 Homo sapi
C 18	35	3.0	164567	3	AC091500	AC091500 Drosophil
C 19	35	3.0	192581	2	PFMAL13P1	AL049180 Plasmodiu
C 20	35	3.0	196194	2	AC104920	AC104920 Mus muscu
C 21	35	3.0	246095	3	AE003802	AE003802 Drosophil
C 22	34	2.9	110000	2	PFMAL4P1_2	Continuation (3 of
C 23	34	2.9	112401	9	AL606462	AL606462 Human DNA
C 24	34	2.9	124170	9	AC006000	AC006000 Homo sapi
C 25	34	2.9	129894	2	AC108181	AC108181 Felis cat
C 26	34	2.9	164008	2	AC115889	AC115889 Mus muscu
C 27	34	2.9	171566	2	AC108651	AC108651 Rattus no
C 28	34	2.9	173916	9	AL449403	AL449403 Human DNA
C 29	34	2.9	175128	2	AC106764	AC106764 Homo sapi
C 30	34	2.9	182507	9	AC013441	AC013441 Homo sapi
C 31	34	2.9	210317	2	AC115074	AC115074 Mus muscu
C 32	34	2.9	219014	2	AC024141	AC024141 Mus muscu
C 33	34	2.9	235085	2	AC094364	AC094364 Rattus no
C 34	34	2.9	247794	2	AC126044	AC126044 Mus muscu
C 35	33	2.8	224	8	AY018093	AY018093 Oryza sat
C 36	33	2.8	224	8	AY018093	AY018093 Oryza sat
C 37	33	2.8	224	8	AY018107	AY018107 Oryza sat
C 38	33	2.8	224	8	AY020625	AY020625 Oryza sat
C 39	33	2.8	224	8	AY020625	AY020625 Oryza sat
C 40	33	2.8	228	8	AY020816	AY020816 Oryza sat
C 41	33	2.8	236	8	AY021075	AY021075 Oryza sat
C 42	33	2.8	634	10	AF401532	AF401532 Mus muscu
C 43	33	2.8	634	10	AF401532	AF401532 Mus muscu
C 44	33	2.8	3089	8	YSCMSL1	J03495 Yeast (S. ce
C 45	33	2.8	4508	8	SCNNA2	X05143 Yeast NAM2

ALIGNMENTS

RESULT 1

SC9958
S.cerevisiae chromosome XIII cosmid 9958.
246729 271257
246729.1 GI:577134
calcineurin; CNA2; CYB2; cytochrome b2; delta element; GAL80;
inosine-5'-monophosphate dehydrogenase; pili; protein phosphatase;
transfer RNA-Tyr.
Saccharomyces cerevisiae.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 20951)
Devlin,K. and Churcher,C.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1
Barrell, B.
Direct Subm
Submitted (C
sequencing
CB10 1RQ by
barrell@san

COMMENT

All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis. Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons/amino-acids is given for each CDS as is the calculated codon adaptation index (CAI). Cosmid 9958 is overlapped at the start of this sequence by cosmid 9745 and at the end of this sequence by cosmid 9827. Bases 7395 to 7399 and 14397 to 14398 are ambiguous.

FEATURES

```

1..20501 "Saccharomyces cerevisiae"
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 9958"
/complement(1..507)
/gene="PIF1"
/complement(<1..507)
/gene="PIF1"
/notes="Y9958.Olc, incomplete PIF1 gene, len: 169, CAT:
0-14"

```

gene

Gene	Gene- PFI	complement(<1. .507)
CDS		

```

/gene="PIF1"
/note="YM9958.01c, incomplete PIF1 gene, len: 169, CAT:
0.14"
/codon_start=1
/product="pif1"
/protein_id="CAA86714.1"
/db_xref="GI:577135"
/db_xref="SWISS-PROT:P07271"
/translaton="MPKWIRSTLNHIIPRPPICSFNSFLLLKNVSHAKLSFMS
FRSNFTIQLQHKPSILSKEDLLDSDSDDEPDCIOLETEKEKKIITDIHKR
DKPMRDKNVYMFINKSDPLSWDMFKPSITQPPILSFSDOSSOKKSRSTGF
LRPALKKS
846..1976

```

CDs

```

/note="Y99958.02, unknown orf, len: 376, CAI: 0.15"
/codon_start=1
/product="unknown"
/protein_id="CAA86715.1"
/db_xref="GI:577136"
/db_xref="SWISS-PROT:P53397"
/translaton="MSYKGTGLAIKNSCELKIANLVLAQSPSRFVDEKLYQSTTTMKI
GQKQSVVLTRDENEILEVAVVDCGNODALKTHLMKYFRLDVSLSKLHFNQVPI
DQAFRAKLSQGGIRIILQAQPEWTLFVSCSNINISRTMNCNLSNFCGLITTDIG
VAVHSPTESELTSRATEAKLBELGFGYRAXYIETARKLVNDKRAENIISDTTYLQS
ICKDAQVEDKREHLNAPISRRKINLEUHLDFMLFKWGSAGVGVFVVDVHVAQYQISAN
KNHUKUEZKRIALPISRRKINLEUHLDFMLFKWGSAGVGVFVVDVHVAQYQISAN
TTGTGTTKRRKWDKMKITEAIVTQKMLKVELSDLUHIKEAKD"
complement(2194..7233)
/note="Y99958.03c, similar to unknown C. elegans gene
Y0L4_CAEEL zk370.4 CEO0396, and E.coli YCHK_SW:YCHK_ECOLI
P37053 and Shigella SW:YCHK_SHIFL P37054; len: 1679, CAI:
0.17"
/codon_start=1
/product="unknown"
/protein_id="CAA86716.1"

```

CDS

```

Complement[21394, ..., 7233]
/notes="YMN958.03c, similar to unknown C. elegans gene
Y04_CAEEL zk370.4 CE00396, and E.coli YCHK SW:YCHK_ECOLI
P37053 and Shigella SW:YCHK_SHIFL P37054; len: 1679, CAI:
0.17"
/codon_start=1
/codon="unknown"
/protein_id="CAA86716.1"
/db_xref="GI:577137"
/db_xref="SWISS-PROT:O04958"
/translation="MSRMNTNTNTGQTNKSLGSPFNSSNWTYRFQTCLTDOI
ISEAQWLSLSEFNFSWYVYFVMGASRMFRGYGLATLSLLRPNRFFKLHHVQF
TSLSFALFALAVIVTIDYIMKRLISQYKRLTPELNTCSNNSANINAASTQ
SANAPPAAGSTTGASSIIVTSKMLKGNENETFLSSYLDOFLSAIKIFYLEKPV
HDLTKMKTKQMDGIBILLNDTIGFAIVETGTLQHYEVDHGDGDTSDHSDTG
LDQDREDEDEDDDDIDNWTKSCSNLLDEDESVGYHLKNGLGNFOLLNATKVP

```

NPLTSLVILNFTLHSMSSYGNRPPSELSSPIDTIVSVNNMPCSEQNFSNTDSMTV
STNSPFGTSPMPLKVARAAYDCTIGIIPPOSFAKATYKPSASHIIOQVLTKLVHY
TFQTAHYDGLGTEIMDETELVNKSIYVELPYTLKFAVTRKEFTVDKSGSADLEPK
KNSSAKSLKPKPKASDGMTOISQPKIANNNANKSNGINTS'NNOHKRSRSSNNAS
RQONPGDLLSNVPLSRDITLSPNTHANNRKNKSGINTS'NNOHKRSRSSNNAS
VHSKFSLSPELNAQLSTPLSDNTSVIHDHIPSVPVILKRVSPRNLPLPT'NFS
AAQETEDSURLAMEAMLTITLVNKSNNMVSSEYIANMSSLSQNLNEMTSRRSPN
ASFELMSPHCTPSDISVASFASPTQPTMLRILPKEYSITSNRHNKSKSQDKKPRAY
KEELTNLMDPEVDKFAAQGLKPFKKGTITVEQNAKGGLFYIISGRVNTTNSS
SVWSSMPEQVSAQSHKGNPHHTHLLYSVSGSGYGVYLSLIGYKSPVNIIVAKS
VDVYGFSSATLERLFQYFLYIRISDLSLTKSLRLLKDHAEWHILRASSETLFS
QDGSANGYVVLNGLRDKOQUSNSNTSPEVETONTIILGELAQGESGEVEVLTA
MNRYSIVARDSELARIPRTELFELALHPSIMIRVSRVAKKIVGDTVPALTDGP
LSIKENEDSLTPKASTSSLSHKPQNTISGTITFRITLITPTEGSLPVEAPAKML
VQATKQVGRITIGLNNRTTLHLGRASDRLSLKQSGFAELEMITYQTVVYISDTPV
KSNTRTCIACDQDILLADARSPAEIGEBKLLNSKXTARTTELILLHPERYEPG
TLKHWLRYRPNVHHHIGFLSUTITLNMEGKHHVNLNGALAMDKLIQTEFSRKQTQ
NISKGLSGLVIAKTEVESRSMKRSKYTPVVRHUKDRLRLALSGQAIGLVGGG
GARGISHLGVIAETEEOGIPVDVIGSTSTGSPVGGVLYAKDYLPVYGRVKKFAGRI
SIWRKLTDLTPWTSYTG'CHENREIKWTFGDTREDFWIOY'YCNSTNTDSVOEJHS
FYAWRYIRASMSLAGLPLPLEENESSMLDGGV'VDMAYVEMNARQCOTI'FADVGS
DGTREMEYDGLSNGWII'FNRNWPSHPNIPNMAIQVRLGVYSAVNALEKRAITPG
VYVYPRTEAXATLDFSKFEET'HYGVGYGRIFLAGLIDDDKMPYIPGSOET'LNQSV
PEFLHRRNSI'
7395..7399
/note="ambiguous sequence, agaag or gaag"
8358..8672
/note="Y9958.04, unknown orf, len: 104, CAI: 0.32"
/codon_start=1
/product="unknown"
/protein_id="CAA86717.1"
/db_xref="GI:5771138"
/db_xref="SWISS-PROT:Q04964"
/translation="MONSDIFYAQRCCQQAQSTLRITV'NAEFRVRVLP'PMWAVPM
LSTONSGSSASASASLEMEKEDEERLNSIDHNNNKGSGGELKSMFNOCKVEEM
DF"
9155..10969
/gene="CNA2"
9155..10969
/gene="CNA3"
/note="Y9958.05, CNA2 of CMP2 gene, protein phosphatase,
P2B2_YEAST P14747, len: 604, CAI: 0.16"
/codon_start=1
/product="calcineurin"
/protein_id="CAA86718.1"
/db_xref="GI:5771139"
/db_xref="SWISS-PROT:P14747"
/translation="HSSDAIRNTEQ'NAAKILIENTERPQSTTTPIDSKASTVAAN
STATESRDLTQ'LUDDRGVTSNRINMKVAPITASHVPTDGHGFPNGIPRHEFLRD
HFKEGKLSQAQAARIVATLATESLKEPNLISVPAT'IVCGDITGOYFOLLKLFYVGG
DPATTSYFLDGYVDRGSGFSEFLYLYSLKFNLDHEWLLRGNHCKRILTSYFTFKN
EMLKHNYDIEYKCESFNPLPALAMNGOYLCVHIGGISPELNSLODINLNRFPETP
SHGLMCDLWADPIEYDEVDKDLTEEDIVNSKTVPHVHKMAPSRDMFVPSVGRG
SYATFYRAACHFLOETGLLSIIRAHQAQAGYRMVKN'YTLGFSPLLTLFESAPNYLDT
YNNKAATLYNNVNNIKQFNTPHYLPDWMFDM'VTSLPVFGKYTEMLVALNLTIC
TELVAKVSRMTSVLREETNKVQFLKDHSNGVPLASPKHAIILDDHRRKALRKH
ILAVKTPSLSDELKLENKRYEYKVMQKVHEHDAKNDK"
9755..9772
/gene="CNA2"
/note="P500125 Serine/threonine specific protein
phosphatases signature"
complement(join(11169..12283,12692..13151))
/note="Y9958.06c, orf len: 524, CAI: 0.48, similar to
inosine-5'-monophosphate dehydrogenases"
/codon_start=1
/product="putative inosine-5'-monophosphate dehydrogenase"
/protein_id="CAA86719.1"
/db_xref="GI:5771140"
/db_xref="SWISS-PROT:P50094"
/translation="MSAAPDTFKAAIEHLKTYSSKDGSLSVQELMDS'TRGGULTYNDLF
VLPLGNSPMSAVSLQETLTKTILTPRPSVMDPTADNAIYVALLAGIGCFIHNN
CTPRBOASPVKWKWENGFIINSP'VISP'VTVGEYVKWAKRFGFSGPVTEQKCPG

misc feature

CDS

```

/codon_start=1
/product=putative inosine-5'-monophosphate dehydrogenase"
/protein_id="CAA86719.1"
/db_xref="GI:577140"
/db_xref="Swiss-Prot:P50094"
/translation="MSAAPDQKALAEHLKTYSSKQGLSVQELMDSTTGGTLTNYDF
VLPLGVNFPSSAVSLQKLTNLTATVSPMDPTVTADMAIYALAGLQGTFTTHH
LTPDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
CTPEQASQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ

```

```
KLVLVTRDIOQLFEDSLVSEVWTKNPVTGIKIGITLKEGNEILKQTKKLLIVDD
NGNLVMSLRDLMKNQNYPLASKSATTQLLCGAAICTIEADKERLLRLEAGLDVV
ILDSSQSNVFLNNIKWIKETFPDLEIIAGNVADREGOANLIIAAGADGLRIGMSGGS
ICITIEVNAKGPQCTAVYVNCOFANQFVPCMDGQVQNGHITIKALALGSSTVVMG
CMLAGTTESPGEFYFKDKRLKAYRGMGSDAMOKTGNKNASTSRFSESDSVLVAQ
GVSGAVDKGSIKKFIPLYINGLQHSODIGCESITSLEKNVQNGVEVRFERTASAQL
EGVHNLSYERKLYN"
/complete(11730..11768)
/notes="PS00487 IMP dehydrogenase/GMP reductase signature"
/complete(12284..12286)
/notes="splice acceptor sequence, cag"
/complete(12396..12402)
/notes="splice branch sequence tactaac"
/complete(12686..12691)
/notes="splice donor sequence gtagtg"
13765..14301
/notes="YM9958.07, unknown orf, len: 178, CAI: 0.15"
/codon_start=1
/product="unknown"
/protein_id="CAA86720.1"
/db_xref="GI:577141"
/db_xref="SWISS-PROT:Q04969"
/translation="MSSAKPINVYSIPELNOALDEALPSVFAFLNVERBSYALLDAKLY
IGYSANVAGLSFFLDKKEPFDQIVTYQKLLVGAFFVLSLLFWFSRFEIKGTIVYVK
RKGKEEIVYTKFEKNEPLYLVELVQKKGSKKELKAKLEVNKVFENESGYLQND
YFKWFSQHNVLDTKKNE"
14397..14398
/notes="ambiguous sequence, ct or cc"
/complete(14508..16283)
/gene="CYB2"
/complete(14508..16283)
/gene="CYB2"
/notes="YM9958.08c, CYB2 gene, CYB2 YEAST P00175,
cytochrome b2 precursor, len: 591, CAI: 0.21"
/codon_start=1
/product="cytochrome b2 precursor"
/protein_id="CAA86721.1"
/db_xref="GI:577142"
/db_xref="SWISS-PROT:P00175"
/translation="MLKYPLKISKNEAAATLRASKTRLNTIRAYSTVPKSKSFEQ
DSRRKTSQWTLRVGAILAATSSVAYLNWHNQIDNEPKLDNMKQKISPAEVAKINKP
89.6%; Score 1038; DB 8; Length 20951;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1158; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 AATGAGCAACCGTGTCAACAAGAGTCTCAAGACGCGCTACTTATCCCCCAAGSACACGT 60
|||||
DB 7809 AATGAGCAACCGTGTCAACAAGAGTCTCAAGACGCGCTACTTATCCCCCAAGGATCACGT 7868
|||||

QY 61 TCCTTCTGCGCAACATCATTGCGGTGCGAAGCTCGCGCGGTCTTCTGACATTGGTAAGAA 120
|||||
DB 7869 TCCTTCTGCGCAACATCATTGCGGTGCGAAGCTCGCGCGGTCTTCTGACATTGGTAAGAA 7928
|||||

QY 121 TACTTCCAACTAAGACATGCTTCTCTTTTGTAGGCCAATGATAGGAAGAACA 180
|||||
DB 7929 TACTTCCAACTAAGACATGCTTCTCTTTTGTAGGCCAATGATAGGAAGAACA 7988
|||||

QY 181 TAGATTATAAATACGTGCAAGATATAGTAGATATGTTTTATGTTAGACCTCGTACATAG 240
|||||
DB 7989 TAGATTATAAATACGTGCAAGATATAGTAGATATGTTTTATGTTAGACCTCGTACATAG 8048
|||||

QY 241 GAATAATTGACGTTTTTTTTTGGCCCAACATTTGAAATTTTTTTTTTACCTCGGCGCTCA 300
|||||
DB 8049 GAATAATTGACGTTTTTTTTTGGCCCAACATTTGAAATTTTTTTTTTACCTCGGCGCTCA 8108
|||||

QY 301 GCCCAACGGGTCCACTACCGCGCGGTGCGCAATTTTGGGAAGTFCATCCGTGCCAAAA 360
|||||
DB 8109 GCCCAACGGGTCCACTACCGCGCGGTGCGCAATTTTGGGAAGTFCATCCGTGCCAAAA 8168
|||||

QY 361 AGSAAATACCCATACATATCGTTACTGTTTGGACATCGCCCGTTCCGCCGATTCG 420
|||||
DB 8169 AGSAAATACCCATACATATCGTTACTGTTTGGACATCGCCCGTTCCGCCGATTCG 8228
|||||
```

```
QY 421 CCTACGCGGTATAAAAAGAGATCTTTTTTCTGGCTGCCCTT-CCATTTTAAAT 479
|||||
DB 8229 CCTACGCGGTATAAAAAGAGATCTTTTTTCTGGCTGCCCTTCCATTTTAAAT 8288
|||||

QY 480 GTCTATTCTGCTCCTTTGTGATCTTACGGTCTCACTAAGCTCTCTTCAACTGCTCAATAA 539
|||||
DB 8289 GTCTATTCTGCTCCTTTGTGATCTTACGGTCTCACTAAGCTCTCTTCAACTGCTCAATAA 8348
|||||

QY 540 TTTCCCGGTATGCAAAATTTCCCAAGACTACTTTTACGCTCAAAATCGCTGCCAACAACA 599
|||||
DB 8349 TTTCCCGGTATGCAAAATTTCCCAAGACTACTTTTACGCTCAAAATCGCTGCCAACAACA 8408
|||||

QY 600 CAAGCCCTTCCACATTCGCTACCGTGACCATGGGGGAATTTAGAGGGTGCCTTTGCCA 659
|||||
DB 8409 CAAGCCCTTCCACATTCGCTACCGTGACCATGGGGGAATTTAGAGGGTGCCTTTGCCA 8468
|||||

QY 660 CTTATGGCTGAGGTTCTTATGTTGCTACTCAAACTCCATGGGAGCTCCGCTTCGCC 719
|||||
DB 8469 CTTATGGCTGAGGTTCTTATGTTGCTACTCAAACTCCATGGGAGCTCCGCTTCGCC 8528
|||||

QY 720 TCCGCTCTTCTTCAATAGAAATGTTGGAAAGAGATTTGGAGGAGAGACTCAACTCTATCGAT 779
|||||
DB 8529 TCCGCTCTTCTTCAATAGAAATGTTGGAAAGAGATTTGGAGGAGAGACTCAACTCTATCGAT 8588
|||||

QY 780 CATGACATGAACAACAATAATTTGGTCTGCGGAATCTATGTTTCAACACAGGT 839
|||||
DB 8589 CATGACATGAACAACAATAATTTGGTCTGCGGAATCTATGTTTCAACACAGGT 8648
|||||

QY 840 AAGTCGAGGAATGAGCTTCTTAAAGTCTCTTTCATACACTCTTTTCTCTTCCATT 899
|||||
DB 8649 AAGTCGAGGAATGAGCTTCTTAAAGTCTCTTTCATACACTCTTTTCTCTTCCATT 8708
|||||

QY 900 TCCCACTAGTCTCTTCTTCTCTTAGATACCTCTTTTTCAGGAGACTCTGCTCT 959
|||||
DB 8709 TCCCACTAGTCTCTTCTTCTCTTAGATACCTCTTTTTCAGGAGACTCTGCTCT 8768
|||||

QY 960 ACTATTGTTGCTCTCTCGAAACATCTCTCCCGTGCATTTTCTCTTCCCTTTATATACA 1019
|||||
DB 8769 ACTATTGTTGCTCTCTCGAAACATCTCTCCCGTGCATTTTCTCTTCCCTTTATATACA 8828
|||||

QY 1020 TATATATATATATATATATATATGCTCTCTACGATTTTGTATTTCTGCTGCTTAT 1079
|||||
DB 8829 TATATATATATATATATATATATGCTCTCTACGATTTTGTATTTCTGCTGCTTAT 8888
|||||

QY 1080 CAAAGATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCAACATTTGCCCC 1139
|||||
DB 8889 CAAAGATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCAACATTTGCCCC 8948
|||||

QY 1140 CTCTCTTGATCAATGCTTT 1158
|||||
DB 8949 CTCTCTTGATCAATGCTTT 8967
|||||

RESULT 2
SCCMP2 Yeast CMP2 gene for calmodulin-binding protein 2.
LOCUS X54964
DEFINITION X54964.1 GI:3559
VERSION calmodulin binding protein; CMP2 gene.
KEYWORDS Saccharomyces cerevisiae.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 2998)
AUTHORS Liu,Y., Ishii,S., Tokai,M., Tsutsumi,H., Ohki,O., Akada,R., Tanaka,K., Tsuchiya,E., Fukui,S. and Miyakawa,T.
TITLE The Saccharomycetes cerevisiae genes (CMP1 and CMP2) encoding calmodulin-binding proteins homologous to the catalytic subunit of mammalian protein phosphatase 2B
JOURNAL Mol. Gen. Genet. 227 (1), 52-59 (1991)
MEDLINE 91260679
```

```

PUBMED 1646387
REFERENCE 2 (bases 1 to 2998)
AUTHORS Miyakawa,T.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1990) Miyakawa T., Department of Fermentation
Technology, Faculty of Engineering, Hiroshima University, Saijo,
Higashi-Hiroshima 724, Japan
COMMENT See X54963 for related sequence.
FEATURES
    source
        1..2998
            /organism="Saccharomyces cerevisiae"
            /strain="A364A"
            /db_xref="taxon:4932"
            /clone_lib="lambda gtl1"
            539..545
                /note="consensus alpha-factor response element"
            663..669
                /note="consensus alpha-factor response element"
            807..813
                /note="consensus alpha-factor response element"
            907..921
                /gene="CMP2"
            907..921
                /gene="CMP2"
            /codon_start=1
            /product="calmodulin-binding protein 2"
            /protein_id="CAA38712.1"
            /db_xref="GI:3560"
            /db_xref="SWISS-PROT:P14747"
            /translation="MSSDAIRNTEQINAAIKIENKTERPOSSTPIDSKASTVAAAAN
STATSTRLDQYTLDDGRVSTNRIMNKVPATISHTVPTDEDVQPNIGIRPHEFLRD
HFREGKLSAAQARIVLATLEFSKPNLISVPAPITVCGDIHQOYFDLLKLFVGG
DPATSYLFDGYDRGSEFSECLIIYLSKLNENDHFWLLRGNHCECHLTSYTFEKN
EHLKNDLIEKCESENNPLAALMGQYLCVHGGSPELNSQDINNLRPREIP
SHGLMDLADPIEEYDEVLDKLTDEEDIVNSKTMVPHGKMAPSRDMFVPSVRGC
SYAFTRAACHFLQETGLLSIRAEADAGYRMVKNKTLGFPSSLTLFSAFNYLDT
YNNKAALIKYNNVNNVIRQFNNTPHYWLPDMFVFTWSLPFVGEKVTEMLVAIINIC
TELELNTPIVEELVGTGDKLPQAGKSEATPOPATSPKPHASTLDEHRRKALRNK
ILAVAKVSRMSVLPREEFNKVOFLKDHNSVLPKALSGVKGDLDEALSTPERARKHD
LINEKLPSLDELKNNKKYKQVQKVEHDAKNDK"
            2407..2475
                /gene="CMP2"
            /bound_moiety="calmodulin"
BASE COUNT 878 a 673 c 558 g 889 t
ORIGIN

Query Match 59.7%; Score 691; DB 8; Length 2998;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 CCATTTTAAATGCTTATCTGCTCCTTTGTGATCTTACGGTCTCACTAAACCTCTCTTCA 527
DB 29 CCATTTTAAATGCTTATCTGCTCCTTTGTGATCTTACGGTCTCACTAAACCTCTCTTCA 88

QY 528 ACTGCTCAATATTTCCGCGTATGCAAAATTTCCCAAGACTACTTTTACGCTCAAAATCGC 587
DB 89 ACTGCTCAATATTTCCGCGTATGCAAAATTTCCCAAGACTACTTTTACGCTCAAAATCGC 148

QY 588 TGCCAAACAACAAGCCCTTCCACATTTGGTACCGTACCATGGCGGAATTTAGAAG 647
DB 149 TGCCAAACAACAAGCCCTTCCACATTTGGTACCGTACCATGGCGGAATTTAGAAG 208

QY 648 GTGCTTTGCCACCTATGCTGAGTCTCTATGTTGTCTACTCAAACTCCATGGCGAGC 707
DB 209 GTGCTTTGCCACCTATGCTGAGTCTCTATGTTGTCTACTCAAACTCCATGGCGAGC 268

QY 708 TCCGCTTCTGCTCGCTCTTCTTATTAGAAATGTGGGAAAAGATTGTGGAGGAGACTC 767
DB 269 TCCGCTTCTGCTCGCTCTTCTTATTAGAAATGTGGGAAAAGATTGTGGAGGAGACTC 328

QY 768 AACTCTATCGATCATGACATGAACAACAACAATTTGGTTCTGGCGAACTAAATCTATG 827
DB 329 AACTCTATCGATCATGACATGAACAACAACAATTTGGTTCTGGCGAACTAAATCTATG 388

```

```

QY 828 TTCACACAGGTAAAGTCGAGGAATGACATCTTAAAGTCTCTTTCATACACTCTTTCTTT 887
DB 389 TTCACACAGGTAAAGTCGAGGAATGACATCTTAAAGTCTCTTTCATACACTCTTTCTTT 448

QY 888 TCTCTTTCCATTTCCCACTAGTTCGTCTTTTCTCTCTTAGATACACCTCTCTTTTCAGG 947
DB 449 TCTCTTTCCATTTCCCACTAGTTCGTCTTTTCTCTCTTAGATACACCTCTCTTTTCAGG 508

QY 948 GACTCTCGTCTACTATTGTTGTCTCATCTCGAACAATCTCTCCCGTGCATTTTCCTTTC 1007
DB 509 GACTCTCGTCTACTATTGTTGTCTCATCTCGAACAATCTCTCCCGTGCATTTTCCTTTC 568

QY 1008 CCTTTATACATATATATATATATATATATATATATATATATATATATATATATATAT 1067
DB 569 CCTTTATACATATATATATATATATATATATATATATATATATATATATATATATATAT 628

QY 1068 CTGTCTCTTTATCAAGATAGTCTATATACCTTTTGATACAGTATAGCTAGCGCC 1127
DB 629 CTGTCTCTTTATCAAGATAGTCTATATACCTTTTGATACAGTATAGCTAGCGCC 688

QY 1128 AACATTGCCCCCTCTCTTGATCAATGCTTT 1158
DB 689 AACATTGCCCCCTCTCTTGATCAATGCTTT 719

RESULT 3
190044
LOCUS I90044 1767 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 20 from patent US 5723436.
ACCESSION I90044
VERSION I90044.1 GI:3409984
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1767)
AUTHORS Huang,L. and Cyert,M.S.
TITLE Calcineurin interacting protein compositions and methods
JOURNAL Patent: US 5723436-A 20 03-MAR-1998;
FEATURES
    Location/Qualifiers
        1..1767
            /organism="unknown"
BASE COUNT 524 a 416 c 353 g 474 t
ORIGIN

Query Match 6.4%; Score 74; DB 6; Length 1767;
Best Local Similarity 100.0%; Pred. No. 2,9e-29;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1085 ATAGTCTATAATACGTTTGTATACAGCTAGATATCGCTAGCGCAACATTTGCCCTCTC 1144
DB 1 ATAGTCTATAATACGTTTGTATACAGCTAGATATCGCTAGCGCAACATTTGCCCTCTC 60

QY 1145 TTGATCAATGCTTT 1158
DB 61 TTGATCAATGCTTT 74

RESULT 4
190040
LOCUS I90040 2353 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 11 from patent US 5723436.
ACCESSION I90040
VERSION I90040.1 GI:3409980
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2353)
AUTHORS Huang,L. and Cyert,M.S.
TITLE Calcineurin interacting protein compositions and methods
JOURNAL Patent: US 5723436-A 11 03-MAR-1998;

```

```

FEATURES
  source      Location/Qualifiers
  1. .2353
  /organism="unknown"
  BASE COUNT  728 a  511 c  461 g  653 t
  ORIGIN
    1085 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 60
    1145 TTGATCAATGCTTT 1158
    1145 TTTGATCAATGCTTT 74

  Query Match      6.4%; Score 74; DB 6; Length 2353;
  Best Local Similarity 100.0%; Pred. No. 2.9e-29;
  Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Qy 1085 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 1144
  Db 1 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 60
  Qy 1145 TTGATCAATGCTTT 1158
  Db 61 TTTGATCAATGCTTT 74

  RESULT 5
  LOCUS      YSCCALA2      2353 bp      DNA      linear      PLN 27-APR-1993
  DEFINITION S.cerevisiae calcineurin A2 (CNA2) gene, complete cds.
  ACCESSION M64840
  VERSION M64840.1 GI:171150
  KEYWORDS  calcineurin A2; calmodulin-binding protein; catalytic subunit;
             phosphoprotein phosphatase.
  SOURCE    Saccharomyces cerevisiae (strain GRF88) (library: Ycp50) DNA.
  ORGANISM  Saccharomyces cerevisiae
             Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
             Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
  REFERENCE 1 (bases 1 to 2353)
             Cyert,M.S., Kunisawa,R., Kaim,D. and Thorner,J.
             Yeast has homologs (CNA1 and CNA2 gene products) of mammalian
             calcineurin, a calmodulin-regulated phosphoprotein phosphatase
             Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7376-7380 (1991)
  JOURNAL   91334468
  MEDLINE   1651503
  PUBLISHED 1651503
  FEATURES
    source      Location/Qualifiers
    1. .2353
    /organism="Saccharomyces cerevisiae"
    /strain="GRF88"
    /db_xref="taxon:4932"
    /cell_type="alpha"
    /tissue_lib="Ycp50"
    262. .2076
    /gene="CNA2"
    262. .2076
    /gene="CNA2"
    /codon_start=1
    /product="calcineurin A2"
    /protein_id="AAA34466.1"
    /db_xref="GI:171151"
    /translation="MSSDAINTTEQINAAIKIENKTERQSSSTPTPIDSASIVAAAN
    STATSRDTQYTLDDGRVYVYNNRNKVPVPAITSHVPTDEELFQPNIGPRHFEFLRD
    HFREGKLSAAQRIYVTLATLFSKEPNLISVPAPITVCGDIHQYFDLLKLFVEVG
    DPATSYFLGDYVDGFSFECILYLSKLNFDHFLIRGNHECKHLTTSYTFKN
    EMLHKYNLDIYEKCESFNPLPLAALMNGOYLCVHGGISPELNSLODINLNLRRETP
    SHGLMCDLLAWDTEYDEVLDKLTDEEDIVNSKTMYPHCKMAPSRDMFVPSVRCG
    SYAFTYAACKHFLQETGLLSIIRAHEAQDAGRYKNTKLGFPFLSLTFSAPNYLDT
    YNNKAALIKYNNNMTRIQFNMTPHYPWLPDMVFTWSLFFVFGKTEMLVAILNIC
    TEDLENDTVIEELVGTDKLQPKGSEATPPQPATSPASPKHASILDDHEHRRALRNK
    ILAVKYSRVSYLVRETNKVOFLKDHNSGVLPRGALSNGVKGLDEALSLSTFERARKHD
    LINKLPPSLDELKLNKYYEKVQWVHEHDAKNDK"
    BASE COUNT  728 a  511 c  461 g  653 t
    ORIGIN      chromosome XIII near the SUP5 locus.

  Query Match      6.4%; Score 74; DB 8; Length 2353;
  Best Local Similarity 100.0%; Pred. No. 2.9e-29;
  Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Qy 1085 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 1144
  Db 1 ATAGTCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTCTC 60
  Qy 1145 TTGATCAATGCTTT 1158
  Db 61 TTTGATCAATGCTTT 74

  RESULT 6
  LOCUS      D38434      2134 bp      DNA      linear      ROD 19-JUN-1999
  DEFINITION Mus musculus NCAM gene exons.
  ACCESSION D38434
  VERSION D38434.1 GI:3219274
  KEYWORDS  alternative splicing; NCAM.
  SOURCE    Mus musculus (strain:Balb/c) Adult Liver DNA, clone:cosmid clone M1
             sub_clone:Clontech (Cat.#ML1037m).
  ORGANISM  Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE 1 (bases 1 to 2134)
             Kawahigashi,H., Harada,Y., Asano,A. and Nakamura,M.
             A cis-acting regulatory element that affects the alternative
             splicing of a muscle-specific exon in the mouse NCAM gene
             Biochim. Biophys. Acta 1397 (3), 305-315 (1998)
  JOURNAL   98250618
  MEDLINE
  REFERENCE 2 (bases 1 to 2134)
             Nakamura,M.
             Direct Submission
             Submitted (29-SEP-1994) Masahiko Nakamura, Osaka University,
             Institute for Protein Research; 3-2 Yamadaoka, Suita, Osaka 565,
             Japan (E-mail:masahiko@protein.osaka-u.ac.jp, Tel:06-879-8628,
             Fax:06-879-8629)
  JOURNAL
  TITLE
  JOURNAL
  FEATURES
    source      Location/Qualifiers
    1. 2134
    /organism="Mus musculus"
    /strain="Balb/c"
    /db_xref="taxon:10090"
    /clone="cosmid clone M1"
    /sub_clone="Clontech (Cat.#ML1037m)"
    /tissue_type="Liver"
    /dev_stage="Adult"
    246. .2024
    /gene="NCAM"
    246. .293
    /gene="NCAM"
    /note="Exon MSDb is subject of alternative splicing
    exon MSDb"
    /evidence=not_experimental
    1983. .2024
    /gene="NCAM"
    /note="Exon MSDC is subjected alternative splicing
    exon MSDC"
    /evidence=not_experimental
    569 a  493 c  441 g  625 t
    BASE COUNT  569 a  493 c  441 g  625 t
    ORIGIN

  Query Match      3.2%; Score 37; DB 10; Length 2134;
  Best Local Similarity 100.0%; Pred. No. 7.4e-09;
  Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Qy 1012 TATATACATATATATATATATATATATATATATATATATATATATATATATATATAT 1048
  Db 1765 TATATACATATATATATATATATATATATATATATATATATATATATATATATATAT 1801

  RESULT 7
  LOCUS      AB001873      2648 bp      DNA      linear      ROD 07-JUL-1998
  DEFINITION Mus musculus gene for neural cell adhesion molecule, specific
             sequence domains b and c, partial cds.
  ACCESSION AB001873
  VERSION AB001873.1 GI:3293066
  KEYWORDS  NCAM MSDb and MSDC; neural cell adhesion molecule; muscle specific

```


corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/chrX>
 RP13-166C10 is from the library RPCI-13.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP13-166C10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP13-166C10 is at 1 in this sequence.
 The true left end of clone RP13-212L9 is at 121950 in this sequence.

FEATURES

```

source      Location/Qualifiers
1..122049
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.1-21.3"
/clone="RP13-166C10"
/clone.lib="RPCI-13.1"
1..209
/note="L1P3 repeat: matches 1..210 of consensus"
213..3460
/note="L1M1 repeat: matches 1227..4605 of consensus"
3462..3859
/note="MLN2B repeat: matches 1..404 of consensus"
3862..3951
/note="45 copies 2 mer at 84% conserved"
3955..4007
/note="MLN2B repeat: matches 394..448 of consensus"
4010..4894
/note="L1M1 repeat: matches 4604..5403 of consensus"
4896..5387
/note="L1P2 repeat: matches 5291..5782 of consensus"
5384..5421
/note="L1P3 repeat: matches 6109..6146 of consensus"
5427..6098
/note="L1P16 repeat: matches 5463..6132 of consensus"
6868..7139
/note="AlusX repeat: matches 3..293 of consensus"
8289..8346
/note="29 copies 2 mer at 75% conserved"
9363..9643
/note="MER83-internal repeat: matches 2433..2739 of consensus"
9509..9761
/note="match: GSS: Em:B60254"
9556..9699
/note="match: GSS: Em:AQ304805"
9569..9760
/note="match: GSS: Em:B50201"
9712..9957
/note="L1P44 repeat: matches 5901..6146 of consensus"
10756..10809
/note="MER53 repeat: matches 124..177 of consensus"
complement(11008..11749)
/note="match: GSS: Em:AQ490113"
11157..11518

```

```

misc_feature
11609..12030
/note="THE1C repeat: matches 1..371 of consensus"
1287..13458
/note="match: GSS: Em:AQ227628"
/note="L1MB8 repeat: matches 4854..6166 of consensus"
14886..14945
/note="L1 repeat: matches 5117..5177 of consensus"
14951..15325
/note="L1MA10 repeat: matches 5959..6322 of consensus"
15518..16736
/note="L1P48 repeat: matches 4936..6159 of consensus"
16811..17157
/note="L1ME repeat: matches 1122..1491 of consensus"
17415..17874
/note="L1M4 repeat: matches 2879..3356 of consensus"
17875..18175
/note="Alusg1 repeat: matches 1..301 of consensus"
18176..18358
/note="L1M4 repeat: matches 3356..3534 of consensus"
18419..18781
/note="MLN2B repeat: matches 1..370 of consensus"
18782..24281
/note="HERVL repeat: matches 116..5654 of consensus"
24282..24673
/note="MLN2B repeat: matches 1..404 of consensus"
24676..24737
/note="31 copies 2 mer at 77% conserved"
24739..24782
/note="MLN2B repeat: matches 406..448 of consensus"
24784..25313
/note="L1M4 repeat: matches 3944..4483 of consensus"
25324..25958
/note="L1P13 repeat: matches -650..-32 of consensus"
26411..28072
/note="L1P16 repeat: matches 751..2405 of consensus"
28084..31548
/note="L1P16 repeat: matches 2555..6066 of consensus"
31561..32389
/note="L1R1 repeat: matches 1..785 of consensus"
32451..32528
/note="L1P16 repeat: matches 6080..6157 of consensus"
32529..33897
/note="L1MD1 repeat: matches 4477..5931 of consensus"
34059..34309
/note="L1ME1 repeat: matches 5921..6168 of consensus"
34330..35459
/note="L1ME2 repeat: matches 5018..6154 of consensus"
35659..35877
/note="L1MEC repeat: matches 2087..2310 of consensus"
35880..36426
/note="match: GSS: Em:AQ171443"
35989..36329
/note="L1R16A repeat: matches 97..450 of consensus"
36716..36880
/note="match: STS: Em:L77831"
37230..37335
/note="L1M4 repeat: matches 3803..3906 of consensus"
37495..37616
/note="MLN2CA repeat: matches 3..125 of consensus"
37887..38153
/note="L1M4 repeat: matches 3988..4250 of consensus"
38437..38498
/note="31 copies 2 mer at 87% conserved"
40355..40737
/note="MSTB repeat: matches 1..426 of consensus"
40738..40915
/note="MSTB-internal repeat: matches 1471..1651 of consensus"
40908..41491
/note="MSTB-internal repeat: matches 743..1327 of consensus"
41498..43173
/note="MER52A repeat: matches 21..1755 of consensus"

```

```

repeat_region 43178..43487
/Note="MSTRB-internal repeat: matches 431..708 of
consensus"
43485..43980
/Note="MSTRB-internal repeat: matches 2..491 of consensus"
43981..44344
/Note="MSTRB repeat: matches 1..426 of consensus"
44345..44719
/Note="MIR repeat: matches 72..188 of consensus"
44720..45044
/Note="MIR repeat: matches 57..254 of consensus"
45045..45435
/Note="MER9A repeat: matches 98..128 of consensus"
45436..45842
/Note="L1PA14 repeat: matches 5492..6149 of consensus"
45843..46161
/Note="AluSq repeat: matches 1..305 of consensus"
46162..46332
/Note="L1PA14 repeat: matches 5322..5492 of consensus"
46333..46617
/Note="AluSq repeat: matches 39..297 of consensus"
46618..46911
/Note="L1PA14 repeat: matches 5029..5322 of consensus"
46912..47210
/Note="AluSq repeat: matches 1..297 of consensus"
47211..47384
/Note="L1PA14 repeat: matches 4856..5029 of consensus"
47385..47533
/Note="L1PA12 repeat: matches 193..3412 of consensus"
47534..47585
/Note="L1MC4 repeat: matches 7703..7977 of consensus"
47586..47982
/Note="L1MC4 repeat: matches 7255..7508 of consensus"

Query Match 3.1%; Score 36; DB 9; Length 122049;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1007 CCCTTTATACATATATATATATATATATATATATATATAT 1042
Db 24656 CCCTTTATACATATATATATATATATATATATATATATAT 24691

RESULT 10
AC099011/c 173380 bp DNA linear INV 08-NOV-2001
LOCUS Drosophila melanogaster, chromosome 2L, region 24C-24D, BAC clone
DEFINITION BACR22M14, complete sequence.
ACCESSION AC099011.1 GI:16798943
VERSION AC099011.1
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 173380)
AUTHORS Celnikier,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE Sequencing of Drosophila chromosome 2L, region 24C-24D
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173380)

repeat_region 43178..43487
/Note="MSTRB-internal repeat: matches 431..708 of
consensus"
43485..43980
/Note="MSTRB-internal repeat: matches 2..491 of consensus"
43981..44344
/Note="MSTRB repeat: matches 1..426 of consensus"
44345..44719
/Note="MIR repeat: matches 72..188 of consensus"
44720..45044
/Note="MIR repeat: matches 57..254 of consensus"
45045..45435
/Note="MER9A repeat: matches 98..128 of consensus"
45436..45842
/Note="L1PA14 repeat: matches 5492..6149 of consensus"
45843..46161
/Note="AluSq repeat: matches 1..305 of consensus"
46162..46332
/Note="L1PA14 repeat: matches 5322..5492 of consensus"
46333..46617
/Note="AluSq repeat: matches 39..297 of consensus"
46618..46911
/Note="L1PA14 repeat: matches 5029..5322 of consensus"
46912..47210
/Note="AluSq repeat: matches 1..297 of consensus"
47211..47384
/Note="L1PA14 repeat: matches 4856..5029 of consensus"
47385..47533
/Note="L1PA12 repeat: matches 193..3412 of consensus"
47534..47585
/Note="L1MC4 repeat: matches 7703..7977 of consensus"
47586..47982
/Note="L1MC4 repeat: matches 7255..7508 of consensus"

Query Match 3.1%; Score 36; DB 9; Length 122049;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1007 CCCTTTATACATATATATATATATATATATATATATATAT 1042
Db 24656 CCCTTTATACATATATATATATATATATATATATATATAT 24691

RESULT 11
AC008003/c 181728 bp DNA linear INV 10-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2L, region 24D-24D, BAC clone
DEFINITION BACR48D03, complete sequence.
ACCESSION AC008003
VERSION AC008003.8 GI:13270546
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 181728)
AUTHORS Celnikier,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE Sequencing of Drosophila chromosome 2L, region 24C-24D
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173380)

```

Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
 Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M.,
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
 Sequencing of Drosophila chromosome 2L, region 24D-24D
 Unpublished
 TITLE JOURNAL
 REFERENCE
 AUTHORS
 2 (bases 1 to 181728)
 Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
 Svirkas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Direct Submission
 Submitted (09-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Mar 10, 2001 this sequence version replaced gi:7264779.
 Sequence submitted by:
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdp@fruitfly.berkeley.edu.
 Location/Qualifiers
 FEATURES
 source
 1. 181728
 /organism="Drosophila melanogaster"
 /strain="y: cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2L"
 /map="24D-24D"
 /clone="BACR48D02 (D851)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACE3.6)"
 BASE COUNT 55122 a 35756 c 36238 g 54612 t
 ORIGIN
 Query Match 3.1% Score 36; DB 3; Length 181728;
 Best Local Similarity 100.0%; Pred. No. 2.9e-08;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1009 CTTTATATACATATATATATATATATATATATATATATGCT 1044
 Db 1097 CTTTATATACATATATATATATATATATATATATATATGCT 1062
 RESULT 12
 AC008327 185469 bp DNA linear HTG 20-AUG-1999
 LOCUS
 Drosophila melanogaster chromosome 2 clone BACR14K04 (D859) RPCI-98
 14 K.4 map 27C-27C strain y; cn bw sp, *** SEQUENCING IN PROGRESS
 *** 188 unordered pieces.
 AC008327
 AC008327.2 GI:5748862
 HTG; HTGS-PHASE1.
 Drosophila melanogaster.
 Drosophila melanogaster.
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 AUTHORS
 1 (bases 1 to 185469)
 Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
 Svirkas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Direct Submission
 Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdp@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 188 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 672: contig of 672 bp in length
 673 752: gap of unknown length
 753 1942: contig of 1190 bp in length
 1943 2022: gap of unknown length
 2023 2902: contig of 880 bp in length
 2903 2982: gap of unknown length
 2983 4320: contig of 1338 bp in length
 4321 4400: gap of unknown length
 4401 5184: contig of 784 bp in length
 5185 5264: gap of unknown length
 5265 6058: contig of 794 bp in length
 6059 6138: gap of unknown length
 6139 6857: contig of 719 bp in length
 6858 6938: gap of unknown length
 6938 8116: contig of 1179 bp in length
 8117 8196: gap of unknown length
 8197 8894: contig of 697 bp in length
 8894 9623: gap of unknown length
 9624 9703: contig of 650 bp in length
 9704 10744: contig of 1041 bp in length
 10745 10824: gap of unknown length
 10825 11514: contig of 690 bp in length
 11515 11594: gap of unknown length
 11595 12380: contig of 786 bp in length
 12381 12460: gap of unknown length
 12461 13282: contig of 822 bp in length
 13283 13362: gap of unknown length
 13363 13938: contig of 576 bp in length
 13939 14018: gap of unknown length
 14019 15240: contig of 1222 bp in length
 15241 15320: gap of unknown length
 15321 16141: contig of 821 bp in length
 16142 16221: gap of unknown length
 16222 16940: contig of 719 bp in length
 16941 17020: gap of unknown length
 17021 17930: contig of 910 bp in length
 17931 18010: gap of unknown length
 18011 17992: contig of 782 bp in length
 18793 18872: gap of unknown length

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
 Svirkas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Sequencing of Drosophila melanogaster
 Unpublished
 TITLE JOURNAL
 REFERENCE
 AUTHORS
 2 (bases 1 to 185469)
 Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
 Svirkas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Direct Submission
 Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Aug 20, 1999 this sequence version replaced gi:5670411.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdp@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 188 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 672: contig of 672 bp in length
 673 752: gap of unknown length
 753 1942: contig of 1190 bp in length
 1943 2022: gap of unknown length
 2023 2902: contig of 880 bp in length
 2903 2982: gap of unknown length
 2983 4320: contig of 1338 bp in length
 4321 4400: gap of unknown length
 4401 5184: contig of 784 bp in length
 5185 5264: gap of unknown length
 5265 6058: contig of 794 bp in length
 6059 6138: gap of unknown length
 6139 6857: contig of 719 bp in length
 6858 6937: gap of unknown length
 6938 8116: contig of 1179 bp in length
 8117 8196: gap of unknown length
 8197 8893: contig of 697 bp in length
 8894 9623: gap of unknown length
 9624 9703: contig of 650 bp in length
 9704 10744: contig of 1041 bp in length
 10745 10824: gap of unknown length
 10825 11514: contig of 690 bp in length
 11515 11594: gap of unknown length
 11595 12380: contig of 786 bp in length
 12381 12460: gap of unknown length
 12461 13282: contig of 822 bp in length
 13283 13362: gap of unknown length
 13363 13938: contig of 576 bp in length
 13939 14018: gap of unknown length
 14019 15240: contig of 1222 bp in length
 15241 15320: gap of unknown length
 15321 16141: contig of 821 bp in length
 16142 16221: gap of unknown length
 16222 16940: contig of 719 bp in length
 16941 17020: gap of unknown length
 17021 17930: contig of 910 bp in length
 17931 18010: gap of unknown length
 18011 17992: contig of 782 bp in length
 18793 18872: gap of unknown length

Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Eubank, J., Evangelista, C.C., Ferraz, C., Ferreira, S., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacieb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, F., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, C.J.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

20196006
 10731132
 PUBMED

REFERENCE
 2 (bases 1 to 307748)

AUTHORS
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.

TITLE
 Direct Submission

JOURNAL
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT
 On Oct 9, 2000 this sequence version replaced gi:7295742.

FEATURES
 Location/Qualifiers
 1. 307748
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2L"
 <9367..>15865
 /gene="CG10019"
 /evidence=not_experimental
 /map="24B1-24B1"
 /db_xref="FLYBASE:FBan0010019"
 /db_xref="FLYBASE:FBgn0031568"
 join(<9367..9681,9766..10498,10792..10964,11265..11406,12281..12454,12697..12998,13091..13246,13400..13561,13673..13844,13909..14390,14473..14636,15445..>15865)
 /gene="CG10019"
 /product="CT28215"
 /evidence=not_experimental
 /db_xref="FLYBASE:FBan0010019"
 /db_xref="FLYBASE:FBgn0031568"
 join(9367..9681,9766..10498,10792..10964,11265..11406,12281..12454,12697..12998,13091..13246,13400..13561,13673..13844,13909..14390,14473..14636,15445..15865)
 /gene="CG10019"
 /note="CG10019 gene product"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAF51067.1"
 /db_xref="GI:7295764"
 /db_xref="FLYBASE:FBan0010019"
 /translation="MVIQCTDCNCSDRNRSVISCLLPGRSQPYKLGSSFGDDME
 SRIPYRRLNNTSALYGCPLLPSSSTNNRHSWLRIPKNPENSPENPROY
 SSIQTAGHPGSGNTDQATKPKDKSTNQRYLCYECRENNQPLAWWSCRLSLPAE
 QTKMSTTPTTRHCHGGVGTOTLPOTSTPTPLPRFPKSRCHDNFGSYATLPTIEHQ
 MTMPMHCRRHGGGGVGVGVGLGAGVGMPVAPPBPYAPNPNNGCCYHY
 NPMPPAHQSHQHDDHHHAQQHPVQAQHHHPHQVGPQCLCDNHYHPPPTQQQHHH
 QQRANLWANGCYFGYEANHHPSLHATTPHPSHTTDEEDSAFGLADREFLHLRNI
 PALRRGTGVDTRIQRYPDGGGWGIIICGVSLVHLITLGLSLGLLWFAVQYVLYHN

gene
 /gene="CG10020"
 /map="24B1-24B2"
 /db_xref="FLYBASE:FBan0010020"
 /db_xref="FLYBASE:FBgn0031569"
 join(21306..21632,21720..21766,22216..22353,22415..22596,22658..22831,23745..23825,23884..24194,24370..24691,24771..26241,26295..26533,26704..26882,27326..28731,28847..29036,29586..29720,30248..30373,30614..31026,31092..31297,31366..31529,31613..31754,31817..>31868)
 /gene="CG10020"
 /product="CT28217"
 /db_xref="FLYBASE:FBan0010020"
 /db_xref="FLYBASE:FBgn0031569"
 join(22234..22353,22415..22596,22658..22831,23745..23825,23884..24194,24370..24691,24771..26241,26295..26533,26704..26882,27326..28731,28847..29036,29586..29720,30248..30373,30614..31026,31092..31297,31366..31529,31613..31754,31817..>31868)
 /note="CG10020 gene product"
 /codon_start=1
 /protein_id="AAF51066.2"
 /db_xref="GI:10727346"
 /db_xref="FLYBASE:FBan0010020"
 /translation="MDRSYDSEDEHTGRRRLRHTLSTELHPHTSVYSGDIREQVCLT
 DQLHSIEORPRRERFGLSRGOTSGFLGGCVGRVPVDENLAAYADPEKPYR
 NYCTDTHELESSYERRQPASILSTGKSGEADLGRTWICQVQNNPNQSDHSS
 SQERLIGKPIIADNMYETLPPQLPQQVQQQSQSTLPTLPPVQLQPCAA
 HHAPILHQPISATNMVGLIIPQLALPLPSPVELIVKFRNMTQATQDAAA
 RROGLGYNTOMIALSPRPHRIKVSOGATONGKILTKLSEIPNGKDGTS
 HYQGSIVPEMIYRTQSDVTPLQTLSDAONNIMYKPPPLDADHSYGLMGLSR
 TRPSPQNVVNVNAAVALNESFYESNLPRACTSHTDYSNLSLRHITES
 ELMTDPLDFSQSLHLPSPSEYCKNDDDEVEEYDEDEHPHETESYSSVCEQAA
 QHRMSMLPMQIDSPFRDRLRQSMYPVQTEKLDIGFGRSFRDRDKVSCFPDEP
 PAKSQPVEVEVRKVELEPDEDEDESEPEDEDEEEDDEDEDEDEDEDEDE
 LYENITPCGRVPTSDAEDIQDKRSQFRKSVSLDDDYAKASETPPTGLRLPSTP
 ASPCRDELLANVTPPSDSLNDLDRHSGLNESQVTVLTAQRDISGSSSNL
 LITFSKRNLLQHQRSVTDALDDEEQSPYIGLTKLILKTPPTTSRPTSTQ
 PMMPPAITAVTSNQIDLSISSLPKRSMVARGSPDARQLMTSGATKORHSDAS
 FLPMGADAPYRSADISECTNTDEYATDTSKRTGKTPPTTSSSTQSPVSTQ
 SSQLEKHAGSSFSASLSMREDLQDEKDKQTLTKAOLKSPICGVAELTRK
 SPHSISSTSSGCPVLGAALKSPAKESLPQTAVASVTOMKVSSEVIPCIVKPKP
 DSISEDERSEVYSSSGYSEPHEDDEEQTRSKARLQEDERKREKTKMLDIEK
 ENMRALTSPIRKTSSTPEQISGTMDSGSPKMRFRKRIQLRKRKSSREDY
 LAAGAVRSRATPTIFGLSSGSGTELLDMSMTGAVAGTTSAVTSAPSSASKL
 PTERSVTPQEVAVVPAIPAKPHSCATPTSLSPKLTSTGTSKTSDFTLKAKS
 IESLSVSPGSDSVFYSEADGNAAGSQSHCHGCKEMEGQOQNTISELAGDSVESI
 FYIQQIVKPPSDPADSPVTKTTRLYKKMDKRFERSERYHGRGRVYKTRQENIRA
 KSEGRIPSLPNTPLVPAGSSCPVLPDTEQSOHIIYKGYHDAGYTRLLDLDLWTO
 LDHOCFRRRRRSTSEKGFHAKYQVILHVRORCTLEMHQKNSQLIASNS
 TPDSTIRRLLEVLKMLLWVARESYISKLSNGAHGVDTKVVVSGSEGFRTHGSK
 PVVAAIPEPTPAGSSGLEVDIIISVNGQVLDKHTTEVVKIAHDGCEKLEFLVART
 IGVLMHECLEPPSOPIFSGYKQSGQAKGAPNSKRWRFSLPDPNCLVLYKTEDD
 SOLPMCAKMTAKVTDLCVDVQKPFPAKADAGEGIPMYVAADSEMANRQLQLPQAA
 SQDNQWLDKSRCLYQSPNSIQRDPCFGLLLKSGRWCGRKRCVLDKADCLFYQDA
 NSKSAFGMACLHGYKVASMANAGKNSFEIVPETKLRIHYFFCTESEMDKKRWISA
 LEYSIDRWIKSG"

mRNA
 /db_xref="FLYBASE:FBan0010020"
 /db_xref="FLYBASE:FBgn0031569"
 join(21306..21632,21720..21766,22216..22353,22415..22596,22658..22831,23745..23825,23884..24194,24370..24691,24771..26241,26295..26533,26704..26882,27326..28731,28847..29036,29586..29720,30248..30373,30614..31026,31092..31297,31366..31529,31613..31754,31817..>31868)
 /gene="CG10020"
 /product="CT28217"
 /db_xref="FLYBASE:FBan0010020"
 /db_xref="FLYBASE:FBgn0031569"
 join(22234..22353,22415..22596,22658..22831,23745..23825,23884..24194,24370..24691,24771..26241,26295..26533,26704..26882,27326..28731,28847..29036,29586..29720,30248..30373,30614..31026,31092..31297,31366..31529,31613..31754,31817..>31868)
 /note="CG10020 gene product"
 /codon_start=1
 /protein_id="AAF51066.2"
 /db_xref="GI:10727346"
 /db_xref="FLYBASE:FBan0010020"
 /translation="MDRSYDSEDEHTGRRRLRHTLSTELHPHTSVYSGDIREQVCLT
 DQLHSIEORPRRERFGLSRGOTSGFLGGCVGRVPVDENLAAYADPEKPYR
 NYCTDTHELESSYERRQPASILSTGKSGEADLGRTWICQVQNNPNQSDHSS
 SQERLIGKPIIADNMYETLPPQLPQQVQQQSQSTLPTLPPVQLQPCAA
 HHAPILHQPISATNMVGLIIPQLALPLPSPVELIVKFRNMTQATQDAAA
 RROGLGYNTOMIALSPRPHRIKVSOGATONGKILTKLSEIPNGKDGTS
 HYQGSIVPEMIYRTQSDVTPLQTLSDAONNIMYKPPPLDADHSYGLMGLSR
 TRPSPQNVVNVNAAVALNESFYESNLPRACTSHTDYSNLSLRHITES
 ELMTDPLDFSQSLHLPSPSEYCKNDDDEVEEYDEDEHPHETESYSSVCEQAA
 QHRMSMLPMQIDSPFRDRLRQSMYPVQTEKLDIGFGRSFRDRDKVSCFPDEP
 PAKSQPVEVEVRKVELEPDEDEDESEPEDEDEEEDDEDEDEDEDEDEDE
 LYENITPCGRVPTSDAEDIQDKRSQFRKSVSLDDDYAKASETPPTGLRLPSTP
 ASPCRDELLANVTPPSDSLNDLDRHSGLNESQVTVLTAQRDISGSSSNL
 LITFSKRNLLQHQRSVTDALDDEEQSPYIGLTKLILKTPPTTSRPTSTQ
 PMMPPAITAVTSNQIDLSISSLPKRSMVARGSPDARQLMTSGATKORHSDAS
 FLPMGADAPYRSADISECTNTDEYATDTSKRTGKTPPTTSSSTQSPVSTQ
 SSQLEKHAGSSFSASLSMREDLQDEKDKQTLTKAOLKSPICGVAELTRK
 SPHSISSTSSGCPVLGAALKSPAKESLPQTAVASVTOMKVSSEVIPCIVKPKP
 DSISEDERSEVYSSSGYSEPHEDDEEQTRSKARLQEDERKREKTKMLDIEK
 ENMRALTSPIRKTSSTPEQISGTMDSGSPKMRFRKRIQLRKRKSSREDY
 LAAGAVRSRATPTIFGLSSGSGTELLDMSMTGAVAGTTSAVTSAPSSASKL
 PTERSVTPQEVAVVPAIPAKPHSCATPTSLSPKLTSTGTSKTSDFTLKAKS
 IESLSVSPGSDSVFYSEADGNAAGSQSHCHGCKEMEGQOQNTISELAGDSVESI
 FYIQQIVKPPSDPADSPVTKTTRLYKKMDKRFERSERYHGRGRVYKTRQENIRA
 KSEGRIPSLPNTPLVPAGSSCPVLPDTEQSOHIIYKGYHDAGYTRLLDLDLWTO
 LDHOCFRRRRRSTSEKGFHAKYQVILHVRORCTLEMHQKNSQLIASNS
 TPDSTIRRLLEVLKMLLWVARESYISKLSNGAHGVDTKVVVSGSEGFRTHGSK
 PVVAAIPEPTPAGSSGLEVDIIISVNGQVLDKHTTEVVKIAHDGCEKLEFLVART
 IGVLMHECLEPPSOPIFSGYKQSGQAKGAPNSKRWRFSLPDPNCLVLYKTEDD
 SOLPMCAKMTAKVTDLCVDVQKPFPAKADAGEGIPMYVAADSEMANRQLQLPQAA
 SQDNQWLDKSRCLYQSPNSIQRDPCFGLLLKSGRWCGRKRCVLDKADCLFYQDA
 NSKSAFGMACLHGYKVASMANAGKNSFEIVPETKLRIHYFFCTESEMDKKRWISA
 LEYSIDRWIKSG"

CDS
 /db_xref="FLYBASE:FBan0010020"
 /db_xref="FLYBASE:FBgn0031569"
 join(22234..22353,22415..22596,22658..22831,23745..23825,23884..24194,24370..24691,24771..26241,26295..26533,26704..26882,27326..28731,28847..29036,29586..29720,30248..30373,30614..31026,31092..31297,31366..31529,31613..31754,31817..>31868)
 /note="CG10020 gene product"
 /codon_start=1
 /protein_id="AAF51066.2"
 /db_xref="GI:10727346"
 /db_xref="FLYBASE:FBan0010020"
 /translation="MDRSYDSEDEHTGRRRLRHTLSTELHPHTSVYSGDIREQVCLT
 DQLHSIEORPRRERFGLSRGOTSGFLGGCVGRVPVDENLAAYADPEKPYR
 NYCTDTHELESSYERRQPASILSTGKSGEADLGRTWICQVQNNPNQSDHSS
 SQERLIGKPIIADNMYETLPPQLPQQVQQQSQSTLPTLPPVQLQPCAA
 HHAPILHQPISATNMVGLIIPQLALPLPSPVELIVKFRNMTQATQDAAA
 RROGLGYNTOMIALSPRPHRIKVSOGATONGKILTKLSEIPNGKDGTS
 HYQGSIVPEMIYRTQSDVTPLQTLSDAONNIMYKPPPLDADHSYGLMGLSR
 TRPSPQNVVNVNAAVALNESFYESNLPRACTSHTDYSNLSLRHITES
 ELMTDPLDFSQSLHLPSPSEYCKNDDDEVEEYDEDEHPHETESYSSVCEQAA
 QHRMSMLPMQIDSPFRDRLRQSMYPVQTEKLDIGFGRSFRDRDKVSCFPDEP
 PAKSQPVEVEVRKVELEPDEDEDESEPEDEDEEEDDEDEDEDEDEDEDE
 LYENITPCGRVPTSDAEDIQDKRSQFRKSVSLDDDYAKASETPPTGLRLPSTP
 ASPCRDELLANVTPPSDSLNDLDRHSGLNESQVTVLTAQRDISGSSSNL
 LITFSKRNLLQHQRSVTDALDDEEQSPYIGLTKLILKTPPTTSRPTSTQ
 PMMPPAITAVTSNQIDLSISSLPKRSMVARGSPDARQLMTSGATKORHSDAS
 FLPMGADAPYRSADISECTNTDEYATDTSKRTGKTPPTTSSSTQSPVSTQ
 SSQLEKHAGSSFSASLSMREDLQDEKDKQTLTKAOLKSPICGVAELTRK
 SPHSISSTSSGCPVLGAALKSPAKESLPQTAVASVTOMKVSSEVIPCIVKPKP
 DSISEDERSEVYSSSGYSEPHEDDEEQTRSKARLQEDERKREKTKMLDIEK
 ENMRALTSPIRKTSSTPEQISGTMDSGSPKMRFRKRIQLRKRKSSREDY
 LAAGAVRSRATPTIFGLSSGSGTELLDMSMTGAVAGTTSAVTSAPSSASKL
 PTERSVTPQEVAVVPAIPAKPHSCATPTSLSPKLTSTGTSKTSDFTLKAKS
 IESLSVSPGSDSVFYSEADGNAAGSQSHCHGCKEMEGQOQNTISELAGDSVESI
 FYIQQIVKPPSDPADSPVTKTTRLYKKMDKRFERSERYHGRGRVYKTRQENIRA
 KSEGRIPSLPNTPLVPAGSSCPVLPDTEQSOHIIYKGYHDAGYTRLLDLDLWTO
 LDHOCFRRRRRSTSEKGFHAKYQVILHVRORCTLEMHQKNSQLIASNS
 TPDSTIRRLLEVLKMLLWVARESYISKLSNGAHGVDTKVVVSGSEGFRTHGSK
 PVVAAIPEPTPAGSSGLEVDIIISVNGQVLDKHTTEVVKIAHDGCEKLEFLVART
 IGVLMHECLEPPSOPIFSGYKQSGQAKGAPNSKRWRFSLPDPNCLVLYKTEDD
 SOLPMCAKMTAKVTDLCVDVQKPFPAKADAGEGIPMYVAADSEMANRQLQLPQAA
 SQDNQWLDKSRCLYQSPNSIQRDPCFGLLLKSGRWCGRKRCVLDKADCLFYQDA
 NSKSAFGMACLHGYKVASMANAGKNSFEIVPETKLRIHYFFCTESEMDKKRWISA
 LEYSIDRWIKSG"

```

gene
<35384..>.47805
/gene="bowl"
/ncbi="CG10021"
/map="24B1-24C1"
/db_xref="FLYBASE:FBan0010021"
/db_xref="FLYBASE:FBgn0004893"
join(35384..35461,36138..36371,43369..44255,45580..47805)
/gene="bowl"
/product="CT9648"
/db_xref="FLYBASE:FBan0010021"
/db_xref="FLYBASE:FBgn0004893"
join(35703..35993,36138..36371,43369..44255,45580..47805)
/gene="bowl"
/product="CT37221"
/db_xref="FLYBASE:FBan0010021"
/db_xref="FLYBASE:FBgn0004893"
join(42955..42993,43369..44255,45580..47805)
/gene="bowl"
/product="CT40018"
/db_xref="FLYBASE:FBan0010021"
/db_xref="FLYBASE:FBgn0004893"
join(43402..44255,45580..46960)
/gene="bowl"
/ncbi="bowl gene product [alt 1]"
/codon_start=1
/protein_id="AAF31063.1"
/db_xref="GI:7295760"
/db_xref="FLYBASE:FBan0010021"
/db_xref="FLYBASE:FBgn0004893"

Query Match 3.1%; Score 36; DB 3; Length 307748;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1009 CTTTATACATATATATATATATATATATATATATATGT 1044
Db 216810 CTTTATACATATATATATATATATATATATATATATGT 216775

RESULT 14
AE003326
LOCUS Drosophila melanogaster genomic scaffold 142000013385598, complete
DEFINITION
ACCESSION AE003326.2 GI:10729843
VERSION
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 5348)
Adams M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.C., Rogers, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Wortman, J.P., Blazer, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
Andrews, F., Binkov, C., Baldwin, D., Ballew, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Bertram, B.P., Bhandari, D., Bolshakov, S., Borkov, D.,
Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,
Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,

```

```

Wei, M.H., Ibegwam, C., Jaitali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattle, B., McInosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Swirskas, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
Weinstock, G.M., Wellenbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
MEDLINE 10731132
REFERENCE 2 (bases 1 to 5348)
AUTHORS Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gi:7298356.
FEATURES
source
1. 5348
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 1689 a 905 c 890 g 1864 t
ORIGIN
Query Match 3.0%; Score 35; DB 3; Length 5348;
Best Local Similarity 100.0%; Pred. No. 9.6e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1010 TTTATACATATATATATATATATATATATATATATGT 1044
Db 2583 TTTATACATATATATATATATATATATATATATATGT 2617

RESULT 15
DMC52C10/c 44855 bp DNA linear INV 10-FEB-1999
LOCUS Drosophila melanogaster cosmid clone 52C10.
DEFINITION
ACCESSION AL035311
VERSION AL035311.1 GI:4200302
KEYWORDS
SOURCE Drosophila melanogaster.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 44855)
Mottier, Cadieu, Dreano, Lelaure and Galibert, F.
Sequencing the distal X chromosome of Drosophila melanogaster
Unpublished
UPR41 CNRS Recombination Genetics, Université de Rennes 1, 2 av
du Pr Leon Bernard, Faculté de Médecine, 35043 RENNES Cedex, France.
2 (bases 1 to 44855)
Benos, P.
Direct Submission
Submitted (09-FEB-1999) European Drosophila Genome Sequencing
Consortium
Clone=52C10; Contig ID=145; Length=44855; Status=Finished Sequence
submitted by Takis Benos, EMBL Outstation - The EBI, Hinxton,
Cambridge, CB10 1SP, U.K.
E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
Sequencing Consortium. For further information see the European
Drosophila Genome Sequencing Consortium's web site:

```



```

tttgctgcgcagataaacagcctttttaacttaagtgcacatcatcagctgttttt
t
ttggcaatttcaatgaatatcatcaaaagtttagctgcgcaatctatgaatcatttttg
c
atatctaaaagatggcagaatgccaaactogtttcagtatctgcgcatgtccggttttt
g
ttttgctttgatcgtgattttttgtttttttttcttatgacacaaagtatttaaaa
t
gggtaaaacaaagcgtgctgttgactaacacataaaggaaaaagcttcaataatcagag
t
tagtgacacaaagtggaacaaaaaaggaaaaatttggccaagttcaaatgogacagatc
c
acagtcacacgcattttacaaaaaaacaaatgaattcatgaagctgtggccgcgtcag
g
tttaaaagaagaagcgtcaaaagaagaagagacacaccattgctgaaaaagaacaaatgc
g
tagaagttgacattgtatcgaaatttaattggaatgaatatgccaatgttgatgcaga
t
gaggcttgccatggtcaattagatgatgatgaattcgtgcgtctctttagttcaagatg
c
aaaaaccagcgataaacgaagaagaagccatagtgatgaagatgtggacgatactgagcgt
c
ctacttttaaggatgggtttgcagcaatttaaggctttaaagtcacatttttatgcgaaa
c
aataatgatgagtttttgcaaaacttgaaattctatggaagacagcgttttaatttac
a
tataaactcagctgtattgcaaaaaaaattactgactatttttaagttagtttttaaa

```

```

Query Match      3.0%; Score 35; DB 3; Length 44855;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1008 CCTTATATACATATATATATATATATATATAT 1042
Db 15389 CCTTATATACATATATATATATATATATATAT 15355

```

Search completed: July 17, 2003, 16:36:26
Job time : 3061 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	104	27.7	104	2	S49803		hypothetical prote
2	43	11.5	68	2	S78729		protein YML058w-a
3	11	2.9	2206	2	G71611		hypothetical prote
4	10	2.7	97	2	B61988		hypothetical prote
5	10	2.7	97	2	H72173		D5L protein - vari
6	10	2.7	112	2	S48472		probable membrane
7	10	2.7	242	2	T18490		hypothetical prote
8	10	2.7	269	2	H71613		probable multiple
9	10	2.7	321	2	T20437		hypothetical prote
10	10	2.7	321	2	T20437		hypothetical prote
11	10	2.7	580	2	T18439		hypothetical prote
12	10	2.7	600	2	T18446		hypothetical prote
13	10	2.7	605	2	A96660		protein F2K11.20 f
14	10	2.7	1129	2	T19779		hypothetical prote

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x H72173 (1-97)

QY 1013 ATATACATATATATATATATATATATAT 1042
 Db 6 Iletyrilleyriletyriletyriletyr 15

RESULT 6

S48472 probable membrane protein YIL100c-a - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae
 C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 05-Nov-1999

R:Bowman, S.; Church, C.
 submitted to the EMBL Data Library, September 1994

A:Reference number: S48455

A:Accession: S48472

A:Molecule type: DNA

A:Residues: 1-112 <BOW>

A:Cross-references: EMBL:Z38125; NID:g558688; PID:g558706; GSPDB:GN00009; MIPS:YIL100c-a

C:Genetics:

A:Gene: MIPS:YIL100c-a

A:Map position: 9L

C:Keywords: transmembrane protein

F:36-52/Domain: transmembrane #status predicted <TM1>

F:56-74/Domain: transmembrane #status predicted <TM2>

Alignment Scores:

Pred. No.:	0.341	Length:	112
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.67%	Indels:	0
DB:	2	Gaps:	0

US-09-814-661A-1 (1-1158) x S48472 (1-112)

QY 1013 ATATACATATATATATATATATATATAT 1042
 Db 60 Iletyrilleyriletyriletyriletyr 69

RESULT 7

T18490

hypothetical protein C0815c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T18490

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18935

A:Accession: T18490

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-242 <LAW>

A:Cross-references: EMBL:Z98551; PIDN:CAB11129.1

C:Genetics:

A:Map position: 3

A:Note: C0815c

C:Superfamily: Plasmodium falciparum hypothetical protein C0815c

Alignment Scores:

Pred. No.:	0.315	Length:	242
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.67%	Indels:	0
DB:	2	Gaps:	0

US-09-814-661A-1 (1-1158) x T18490 (1-242)

QY 1013 ATATACATATATATATATATATATATAT 1042
 Db 121 Iletyrilleyriletyriletyriletyr 130

RESULT 8

H71613

probable multiple transmembrane domain protein PFB0475c - malaria parasite (Plasmodium falciparum)

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: H71613

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.

; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: H71613

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-269 <GAR>

A:Cross-references: GB:AE001397; GB:AE001362; NID:g3845192; PIDN:AAC71884.1; PID:g384

A:Experimental source: clone 3d7

C:Genetics:

A:Gene: PFB0475c

Alignment Scores:

Pred. No.:	0.311	Length:	269
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.70%	Indels:	0
DB:	2	Gaps:	0

US-09-814-661A-1 (1-1158) x H71613 (1-269)

QY 1041 TATATATATATATATATATATATATATATA 1012
 Db 147 Tyrilleyriletyriletyriletyrile 156

RESULT 9

T20437

hypothetical protein E03G2.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T20437

R:McMurray, A.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19275

A:Accession: T20437

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-321 <WIL>

A:Cross-references: EMBL:Z68113; PIDN:CAA92151.1; GSPDB:GN000028; CESP:E03G2.1

A:Experimental source: clone E03G2

C:Genetics:

A:Gene: CESP:E03G2.1

A:Map position: X

A:Introns: 39/1; 67/1; 122/1; 134/1; 165/1; 193/1; 250/1; 262/2; 298/2

Alignment Scores:

Pred. No.:	0.306	Length:	321
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.67%	Indels:	0
DB:	2	Gaps:	0

US-09-814-661A-1 (1-1158) x T20437 (1-321)

QY 1016 TACATATATATATATATATATATATGTC 1045
 Db 85 Tyrilleyriletyriletyriletyrval 94

RESULT 10

```

submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18446
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-600 <LAW>
A:Cross-references: EMBL:Z98547; NID:el325376; PIDN:CAB51588.1
C:Genetics:
A:Map position: 3
A:Introns: 437/3; 477/2; 502/3; 532/3
A>Note: MAL3P3.11

Alignment Scores:
Pred. No.: 0.286 Length: 600
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.67% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x T18446 (1-600)

QY 1013 ATATACATATATATATATATATATAT 1042
|||||
DB 576 IIEtyrIleYrIleYrIleYrIleYr 585

RESULT 13
A96660
C:Protein F2K11.20 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change
C:Accession: A96660
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Foderspistel, N.A.; Karsan, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizart, L.;
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.;
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.A.;
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana (mouse-ear cress)
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96660
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-605 <STO>
A:Cross-references: GB:AE005173; NID:g6633846; PIDN:AAF19705.1
C:Genetics:
A:Gene: F2K11.20
A:Map position: 1

```

```

Alignment Scores:
Pred. No.:      0.286      Length:      605
Score:          10.00      Matches:      10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      2.67%      Indels:      0
DB:              2          Gaps:      0

US-09-814-661A-1 (1-1158) x A96660 (1-605)

QY      1013 ATATACATATATATATATATATATAT 1042
      |||||
Db      380 lletyRlletyRlletyRlletyR 389

RESULT 14
T18779
hypothetical protein C36B1.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18779
R:Lennard, N.

```

```

submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18446
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-600 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PIDN:CAB51588.1
C:Genetics:
A:Map position: 3
A:Introns: 437/3; 477/2; 502/3; 532/3
A:Note: MAL3P3.11

Alignment Scores:
Pred. No.: 0.286 Length: 600
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.67% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x T18446 (1-600)

QY 1013 ATATACATATATATATATATATATAT 1042
|||||
DB 576 ILETRYILETRYILETRYILETRYILETRY 585

RESULT 13
A96660
protein F2K11.20 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96660
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
chin, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.; Li, J.H.; Liu, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-605 <STO>
A:Cross-references: GB:AE005173; NID:g6633846; PIDN:AAF19705.1; GSPDB:GN00141
C:Genetics:
A:Gene: F2K11.20
A:Map position: 1

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 12:03:25 ; Search time 19.5 Seconds
(without alignments)
4926.112 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 375

Sequence: 1 aatgagcaacgtgtcaaca.....cctctctgtatcaatgcttt 1158

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xl
-O=/cnp2.1/USPTO.spool/US09814661/runat_16072003.115330.15194/app_query.fasta_1.1351
-DB=SwissProt_40 -QPM=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn.1.1.28._runat_16072003.115330.15194 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	27.7	104	1 YMF8_YEAST	Q04964 saccharomyc
2	9	2.4	123	1 YMU7_YEAST	Q04674 saccharomyc
3	8	2.1	101	1 YK27_YEAST	P36140 saccharomyc
4	8	2.2	117	1 YK0_YEAST	P40490 saccharomyc
5	8	2.2	123	1 YMU7_YEAST	Q04674 saccharomyc
6	8	2.1	164	1 YCE1_YEAST	P25571 saccharomyc
7	8	2.2	164	1 YCE1_YEAST	P25571 saccharomyc
8	8	2.2	256	1 RK21_SPIO	P24613 spinacia ol
9	8	2.2	263	1 PYRF_RHIPU	Q94720 rhizomucor
10	8	2.2	264	1 PYRF_SCHPO	P14965 schizosach
11	8	2.2	265	1 PYRF_RHINI	P43230 rhizopus ni
12	8	2.2	265	1 PYRF_RHIRA	P32431 rhizomucor
13	8	2.2	267	1 PYRF_PHYBL	P21593 phycomyces
14	8	2.1	274	1 PSBS_TOBAC	Q95mb4 nicotiana t
15	8	2.2	278	1 PYRF_SCHCO	P14964 schizophyll
16	8	2.2	298	1 PYRF_USTMA	P15188 ustilago ma
17	8	2.1	317	1 WR15_ARATH	O22176 arabidopsis
18	8	2.2	411	1 ALAT_RAT	P17475 rattus norv

19	8	2.1	411	1 CRFL_YARLI	P45815 yarrowia li
20	8	2.2	416	1 ALAT_BOVIN	P34955 bos taurus
21	8	2.2	416	1 ALAT_SHEEP	P12725 ovis aries
22	8	2.2	461	1 PYR5_TOBAC	Q42942 nicotiana t
23	8	2.1	474	1 SOX4_HUMAN	Q06945 homo sapien
24	8	2.2	476	1 PYR5_ARATH	Q42586 arabidopsis
25	8	2.2	486	1 ALIN_ALLSA	Q9cyp7 allium sati
26	8	2.1	492	1 SES3_MOUSE	Q9cyp7 mus musculu
27	8	2.1	522	1 RECN_HELPU	Q94180 helicobacte
28	8	2.1	523	1 C756_CAMME	O04773 campanula m
29	8	2.1	524	1 RECN_HELPU	O25943 helicobacte
30	8	2.1	548	1 GIP2_YEAST	P40036 saccharomyc
31	8	2.2	548	1 NFL_PIG	P02547 sus scrofa
32	8	2.2	554	1 NFL_BOVIN	P02548 bos taurus
33	8	2.2	602	1 DNLI_AERPE	O9vdl8 aeropyrum p
34	8	2.1	670	1 DOT6_YEAST	P40059 saccharomyc
35	8	2.1	787	1 NADE_DROME	O9vya0 drosophila
36	8	2.1	810	1 GARN_DROME	P54362 drosophila
37	8	2.2	810	1 GARN_DROME	P54362 drosophila
38	8	2.2	874	1 SLAP_BACLI	P49052 bacillus li
39	8	2.1	986	1 EPA4_MOUSE	Q03137 mus musculu
40	8	2.1	1018	1 CONT_HUMAN	Q12860 homo sapien
41	8	2.2	1159	1 Y196_HUMAN	Q12768 homo sapien
42	8	2.2	1169	1 SMC_METJA	Q05037 methanococc
43	8	2.1	1251	1 Y0U3_CAEEL	Q09550 caenorhabdi
44	8	2.1	1838	1 CAL5_HUMAN	P20908 homo sapien
45	8	2.2	3829	1 SACS_HUMAN	Q9nzz4 homo sapien

ALIGNMENTS

RESULT 1

YMF8_YEAST
ID YMF8_YEAST STANDARD; PRT; 104 AA.
AC Q04964;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 11.8 kDa protein in OGG1-CNA2 intergenic region.
GN YML058W OR YML958.04
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Devlin K., Church C., Barrell B.G., Rajandream M.A.;
RL Submitted (Nov-1994) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 246729; CAA86717.1; -
DR SGD; S0004523; YML058W.
KW Hypothetical protein.
SQ SEQUENCE 104 AA; 11834 MW; 02735C9178E431B2 CRC64;

Alignment Scores:
Pred. No.: 3.9e-99 Length: 104
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.73% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x YMF8_YEAST (1-104)

2

2

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; 238125; CAA86281.1; -
DR PIR; S48473; S48473.
DR SGD; S0001362; YIL100W.
KW Hypothetical protein.
SQ SEQUENCE 117 AA; 13493 MW; 068281DDE850EF98 CRC64;

Alignment Scores:
Pred. No.: 16.1 Length: 117
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x YIK0_YEAST (1-117)

QY 1036 ATATATATATATATATATATATATAT 1013
Db 33 ILeTyRIleTyRIleTyValTyR 40

RESULT 5

YMU07_YEAST
ID YMU07_YEAST STANDARD; PRT: 123 AA.
AC Q04674;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Very hypothetical 14.3 kDa protein in AAC1-FET3 intergenic region.
GN YMR057C OR YMR796.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; 249703; CAA89767.1; -
DR SGD; S0004661; YMR057C.
KW Hypothetical protein.
SQ SEQUENCE 123 AA; 14292 MW; 36E1A59E7FD67855 CRC64;

Alignment Scores:
Pred. No.: 16 Length: 123
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x YMU7_YEAST (1-123)

QY 1044 ACATATATATATATATATATATATATAT 1021
Db 6 ThrTyRIleTyRIleTyRIleTyR 13

RESULT 6

YCEL_YEAST
ID YCEL_YEAST STANDARD; PRT: 164 AA.
AC P25571;

DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 19.1 kDa protein in PDII-GLK1 intergenic region.
GN YCL041C OR YCL41C.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Dubois E., Pierard A., Gigot D., Glansdorff N., Messenguy F.,
RA Scherens B.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X59720; CAA42374.1; -
DR PIR; S19370; S19370.
DR SGD; S0000546; YCL041C.
KW Hypothetical protein.
SQ SEQUENCE 164 AA; 19050 MW; F84B04923538B702 CRC64;

Alignment Scores:

Pred. No.: 15.2 Length: 164
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.13% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x YCEL_YEAST (1-164)

QY 1017 ACATATATATATATATATATATATATAT 1040
Db 94 ThrTyRIleTyRIleTyRIleTyR 101

RESULT 7

YCEL_YEAST
ID YCEL_YEAST STANDARD; PRT: 164 AA.
AC P25571;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 19.1 kDa protein in PDII-GLK1 intergenic region.
GN YCL041C OR YCL41C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Dubois E., Pierard A., Gigot D., Glansdorff N., Messenguy F.,
RA Scherens B.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X59720; CAA42374.1; -
DR PIR; S19370; S19370.
DR SGD; S0000546; YCL041C.

KW Hypothetical protein.
SQ SEQUENCE 164 AA; 19050 MW; F84B04923538B702 CRC64;

Alignment Scores:
Pred. No.: 15.2 Length: 164
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x YCE1_YEAST (1-164)

QY 1044 ACATATATATATATATATATAT 1021
DB 94 ThrTyrlleTyrlleTyrlleTyr 101

RESULT 8

ID RK21_SPIOL STANDARD; PRT; 256 AA..
 AC P24613;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 50S ribosomal protein L21, chloroplast precursor (CL21) (CS-L7).
 GN RPL21.
 OS Spinacia oleracea (Spinach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
 OX NCBI_TaxID=3562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-67.
 RX MEDLINE=911168304; PubMed=2076556;
 RA Martin W., Lagrange T., Li Y.F., Bisanz-Seyer C., Mache R.;
 RT "Hypothesis for the evolutionary origin of the chloroplast ribosomal
 protein L21 of spinach."
 RL Curr. Genet. 18:553-556(1990).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-75.
 RC STRAIN=cv. Alvaro;
 RX MEDLINE=90375547; PubMed=2398071;
 RA Smooker P.M., Kluft V., Subramanian A.R.;
 RT "A ribosomal protein is encoded in the chloroplast DNA in a lower
 plant but in the nucleus in angiosperms. Isolation of the spinach L21
 protein and cDNA clone with transit and an unusual repeat sequence."
 RL J. Biol. Chem. 265:16699-16703(1990).
 RN [3]
 RP SEQUENCE FROM N.A..
 RC STRAIN=cv. Geant d'hiver;
 RX MEDLINE=93205007; PubMed=8455634;
 RA Lagrange T., Franzetti B., Axelos M., Mache R., Lerbs-Mache S.;
 RT "Structure and expression of the nuclear gene coding for the
 chloroplast ribosomal protein L21: developmental regulation of a
 housekeeping gene by alternative promoters."
 RL Mol. Cell. Biol. 13:2614-2622(1993).
 CC -!- FUNCTION: THIS PROTEIN BINDS TO 23S RIBOSOMAL RNA IN THE PRESENCE
 OF PROTEIN L20 (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE L21P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X56691; CAA40019.1; -
 DR EMBL; M57413; AAA34041.1; -
 DR EMBL; M64682; AAA74715.1; -
 DR PIR; S13527; S13527.
 DR PIR; A48103; A48103.

DR InterPro; IPR001787; Ribosomal_L21p.
 DR Pfam; PF00829; Ribosomal_L21p; 1.
 DR ProDom; PD003604; Ribosomal_L21p; 1.
 DR TIGRFAMs; TIGR00061; L21; 1.
 DR PROSITE; PS01169; rRNA-binding; Chloroplast; Transit peptide.
 KW Ribosomal protein; rRNA-binding; 55 CHLOROPLAST.
 FT TRANSIT 1 55
 FT CHAIN 56 256 50S RIBOSOMAL PROTEIN L21.
 FT DOMAIN 236 253 ALA/GLU-RICH.
 SQ SEQUENCE 256 AA; 28408 MW; 3FD2BE241C42BC79 CRC64;

Alignment Scores:
 Pred. No.: 14.1 Length: 256
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.16% Indels: 0
 DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x RK21_SPIOL (1-256)

QY 730 GAAGAAGCGAGGAGGAGGAGGAG 707
 DB 243 GluGluAlaGluAlaGluAlaGlu 250

RESULT 9

PYRE_RHIPU
 ID PYRE_RHIPU STANDARD; PRT; 263 AA..
 AC Q9720;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
 DE (OMPDCase) (OMPDCase) (Uridine 5'-monophosphate synthase) (UMP
 synthase).
 GN PYR4.
 OS Rhizomucor pusillus.
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
 OC Rhizomucor.
 OX NCBI_TaxID=4840;
 RN [1]
 RP SEQUENCE FROM N.A..
 RC STRAIN=F27;
 RA Yamazaki H., Ohnishi Y., Horinouchi S.;
 RT "Genetic transformation of a Rhizomucor pusillus mutant defective in
 asparagine-linked glycosylation: overproduction of a milk-clotting
 enzyme."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate -> UMP + CO(2).
 CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
 CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB019045; BAA76616.1; -
 CC HSSP; P03962; 1DQW.
 DR InterPro; IPR001754; OMPdecase.
 DR Pfam; PF00215; OMPdecase; 1.
 DR PROSITE; PS00156; OMPDECASE; 1.
 KW Pyrimidine biosynthesis; Lyase; Decarboxylase.
 FT ACT_SITE 93 93 BY SIMILARITY.
 SQ SEQUENCE 263 AA; 29354 MW; 6C5D1E999C437BAE CRC64;

Alignment Scores:
 Pred. No.: 14 Length: 263
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661a-1 (1-1158) x PYRF_RHIPU (1-263)

Qy 611 GGAAGGGCTGTGTTGTTGGCA 588
Db 141 GlyArgGlyLeuLeuLeuAla 148

RESULT 10
PYRF_SCHPO
ID PYRF_SCHPO STANDARD; PRT; 264 AA.
AC P14965;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
DE (OMPDCase) (OMPDecase) (Uridine 5'-monophosphate synthase) (UMP
DE synthase).
GN URA4 OR SPC330.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
RX SEQUENCE FROM N.A.
RP MEDLINE=89201250; PubMed=3241624;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentsley S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goifeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Domineque A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).
CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D17362; BRA04179.1; -
DR HSSP; P03962; IDOW.
DR InterPro; IPR001754; OMPDecase.
DR Pfam; PF00215; OMPDecase; 1.
DR PROSITE; PS00156; OMPDecase; 1.
KW Pyrimidine biosynthesis; Lyase; Decarboxylase.
FT ACT_SITE 93 93 BY SIMILARITY.
SQ SEQUENCE 265 AA; 29611 MW; 87FE1AA5900494FE CRC64;

Alignment Scores:

Pred No.: 14 Length: 264
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661a-1 (1-1158) x PYRF_SCHPO (1-264)

Qy 611 GGAAGGGCTGTGTTGTTGGCA 588
Db 143 GlyArgGlyLeuLeuLeuAla 150

RESULT 11
PYRF_RHINI
ID PYRF_RHINI STANDARD; PRT; 265 AA.
AC P43230;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
DE (OMPDCase) (OMPDecase) (Uridine 5'-monophosphate synthase) (UMP
DE synthase).
GN PYR4.
OS Rhizopus niveus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Rhizopus.
ON NCBI_TaxID=4844;
RX SEQUENCE FROM N.A.
RP STRAIN=Yamazaki / IFO 4810;
RX MEDLINE=96059353; PubMed=7586035;
RA Horuchi H., Takaya N., Yanai K., Nakamura M., Ohta A., Takagi M.;
RT "Cloning of the Rhizopus niveus pyr4 gene and its use for the
RT transformation of Rhizopus delemar.";
RL Curr. Genet. 27:472-478(1995).
CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).
CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D17362; BRA04179.1; -
DR HSSP; P03962; IDOW.
DR InterPro; IPR001754; OMPDecase.
DR Pfam; PF00215; OMPDecase; 1.
DR PROSITE; PS00156; OMPDecase; 1.
KW Pyrimidine biosynthesis; Lyase; Decarboxylase.
FT ACT_SITE 93 93 BY SIMILARITY.
SQ SEQUENCE 265 AA; 29611 MW; 87FE1AA5900494FE CRC64;

Alignment Scores:
```

Pred. No.: 14 Length: 265
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.16% Indels: 0
 DB: 1 Gaps: 0

US-09-814-661a-1 (1-1158) x PYRF_RHINI (1-265)

QY 611 GGAAGGGCTGTGTTGTTGGCA 588

Db 141 GlyArgGlyLeuLeuLeuAla 148

RESULT 12

PYRF_RHIRA STANDARD; PRT: 265 AA.

AC P32431.

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)

DE (OMPDecase) (OMPDecase) (Uridine 5'-monophosphate synthase) (UMP

DE synthase).

GN Rhizomucor racemosus (Mucor circinelloides f. lusitanicus).

OS Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;

OC Mucor.

OX NCBI_TaxID=4841;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92331950; PubMed=1628845;

RA Benito E.P., Diaz-Minguez J.M., Iturriaga E.A., Campuzano V.,

RA Eslava A.P.;

RT "Cloning and sequence analysis of the Mucor circinelloides pyrG gene

RT encoding orotidine-5'-monophosphate decarboxylase: use of pyrG for

RT homologous transformation."

RL Gene 116:59-67(1992).

CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).

CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.

CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M69112; -; NOT_ANNOTATED_CDS.

DR PIR; JC1177; JC1177.

DR HSSP; P03962; LDQW.

DR InterPro; IPR001754; OMPDecase.

DR Pfam; PF00215; OMPDecase; 1.

DR PROSITE; PS00156; OMPDecase; 1.

KW Pyrimidine biosynthesis; Lyase; Decarboxylase.

FT ACT_SITE 93 BY SIMILARITY.

SQ SEQUENCE 265 AA; 29514 MW; 10BF511627BCEA14 CRC64;

Alignment Scores:

Pred. No.: 14 Length: 265
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.16% Indels: 0
 DB: 1 Gaps: 0

US-09-814-661a-1 (1-1158) x PYRF_RHIRA (1-265)

QY 611 GGAAGGGCTGTGTTGTTGGCA 588

Db 141 GlyArgGlyLeuLeuLeuAla 148

RESULT 13

PYRF_PHYBL

ID PYRF_PHYBL STANDARD; PRT: 267 AA.

AC P21593;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)

DE (OMPDecase) (OMPDecase) (Uridine 5'-monophosphate synthase) (UMP

DE synthase).

GN PYRG.

OS Phycomyces blakesleeanus.

OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;

OC Phycomyces.

OX NCBI_TaxID=4837;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL 1555;

RX MEDLINE=91117195; PubMed=2277645;

RA Diaz-Minguez J.M., Iturriaga E.A., Benito E.P., Corrochano L.,

RA Eslava A.P.;

RT "Isolation and molecular analysis of the orotidine-5'-phosphate

RT decarboxylase gene (pyrG) of Phycomyces blakesleeanus."

RL Mol. Gen. Genet. 224:269-278(1990).

CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).

CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.

CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X53601; CAA37670.1; -.

DR PIR; S13736; DCUMOP.

DR HSSP; P03962; LDQW.

DR InterPro; IPR001754; OMPDecase.

DR Pfam; PF00215; OMPDecase; 1.

DR PROSITE; PS00156; OMPDecase; 1.

KW Pyrimidine biosynthesis; Lyase; Decarboxylase.

FT ACT_SITE 95 BY SIMILARITY.

SQ SEQUENCE 267 AA; 29899 MW; CE5B132427F3FDDC CRC64;

Alignment Scores:

Pred. No.: 14 Length: 267
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.16% Indels: 0
 DB: 1 Gaps: 0

US-09-814-661a-1 (1-1158) x PYRF_PHYBL (1-267)

QY 611 GGAAGGGCTGTGTTGTTGGCA 588

Db 143 GlyArgGlyLeuLeuLeuAla 150

RESULT 14

PSBS_TOBAC

ID PSBS_TOBAC STANDARD; PRT: 274 AA.

AC Q9SMB4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Photosystem II 22 kDa protein, chloroplast precursor (CP22).

DE PSBS.

GN Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxID=4097;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. SRL1 TISSUE=Leaf;
 RA Kim S., Pichersky E.;
 RT "Nucleotide sequence of a tobacco psbS gene.";
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Seems to be involved in nonphotochemical quenching, a
 CC process maintains the balance between dissipation and utilization
 CC of light energy to minimize generation of oxidizing molecules,
 CC thereby protecting the plant against photo-oxidative damage (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CHLOROPLAST
 CC THYLAKOID (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ELIP/PSBS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X84225; CAAS9007.1; -
 KW Photosynthesis; Photosystem II; Chloroplast; Transit peptide;
 KW Transmembrane; Repeat.
 FT TRANSIT 1 66 CHLOROPLAST (POTENTIAL).
 FT CHAIN 67 274 PHOTOSYSTEM II 22 KDA PROTEIN.
 FT REPEAT 61 166 1.
 FT REPEAT 167 274 2.
 FT TRANSMEM 104 124 POTENTIAL.
 FT TRANSMEM 138 158 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 SQ SEQUENCE 274 AA; 29087 MW; 591E103D4C33F282 CRC64;

Alignment Scores:
 Pred. No.: 13.9 Length: 274
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.13% Indels: 0
 DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x PSBS_TOBAC (1-274)
 QY 708 TCCGCTTCGCCCTCCGCTCTTCA 731
 Db 44 SerAlaSerAlaSerAlaSer 51
 |||||||||||||||||||

RESULT 15
 PYRF_SCHCO
 ID PYRF_SCHCO STANDARD; PRT; 278 AA.
 AC P14964;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
 DE (OMPDCase) (OMPDecase) (Uridine 5'-monophosphate synthase) (UMP
 DE synthase).
 GN URAL.
 OS Schizophyllum commune (Bracket fungus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Schizophyllaceae; Schizophyllum.
 OX NCBI_TaxID=5334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90060853; PubMed=2684794;
 RA Froeliger E.H., Ullrich R.C., Novotny C.P.;
 RT "Sequence analysis of the URAL gene encoding
 RT orotidine-5'-monophosphate decarboxylase of Schizophyllum commune.";
 RL Gene 83:387-393(1989).

CC -1- CATALYTIC ACTIVITY: Orotidine 5'-phosphate -> UMP + CO(2).
 CC -1- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
 CC -1- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M26019; AAA33928.1; -
 DR PIR; JQ0122; DCSJOS.
 DR HSSP; P03962; 1DQW.
 DR InterPro; IPR001754; OMPDecase.
 DR Pfam; PF00215; OMPDecase; 1.
 DR PROSITE; PS00156; OMPDecase; 1.
 KW Pyrimidine biosynthesis; Lyase; Decarboxylase.
 FT ACT_SITE 95 95 BY SIMILARITY.
 SQ SEQUENCE 278 AA; 29979 MW; 847543EF68288C12 CRC64;

Alignment Scores:
 Pred. No.: 13.9 Length: 278
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.16% Indels: 0
 DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x PYRF_SCHCO (1-278)

QY 611 GGAAGGGCTTGTGTTGTTGCA 588
 Db 143 GlyArgGlyLeuLeuLeuAla 150
 |||||||||||||||||||

Search completed: July 16, 2003, 12:21:12
 Job time : 23.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 17:05:54 ; Search time 1314 Seconds
(without alignments)
2303.418 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: - 537
Sequence: 1 MONSDYFYAQRNQQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09814661/runat_16072003_115309_14272/app_query.fasta_1.263
-DB=GenEmbl -QFT=fastap -SUFFIX=rge -MINMATCH=0.1 -DOOPCL=0 -LOOPEXT=0
-UNITS=blcs -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn_1_1.3745@runat_16072003_115309_14272 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :
1: gb_ba : *
2: gb_htg : *
3: gb_in : *
4: gb_om : *
5: gb_ov : *
6: gb_pat : *
7: gb_ph : *
8: gb_pl : *
9: gb_pr : *
10: gb_pro : *
11: gb_sts : *
12: gb_sy : *
13: gb_un : *
14: gb_vi : *
15: em_ba : *
16: em_fun : *
17: em_hum : *
18: em_in : *
19: em_mu : *
20: em_mu : *
21: em_or : *
22: em_ov : *
23: em_pat : *
24: em_ph : *
25: em_pl : *
26: em_ro : *
27: em_sts : *
28: em_un : *

29: em_vi : *
30: em_htg_hum : *
31: em_htg_inv : *
32: em_htg_other : *
33: em_htg_mus : *
34: em_htg_pln : *
35: em_htg_rod : *
36: em_htg_mam : *
37: em_htg_vrt : *
38: em_sy : *
39: em_htgo_hum : *
40: em_htgo_mus : *
41: em_htgo_other : *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	537	100.0	2998	8	SCCMP2
2	537	100.0	20951	8	SC9958
3	86	16.0	996	9	HSA342553
4	86	16.0	140152	9	HSJ963522
5	82	15.3	8019	3	AF289034
6	82	15.3	8328	3	AF293971
7	82	15.3	163403	3	AC007765
8	82	15.3	164920	3	AC092187
9	82	15.3	306436	3	AE003581
10	81	15.1	231560	2	AC127287
11	81	15.1	300895	2	AC079431
12	80.5	15.0	33089	3	AF016428
13	80.5	15.0	71607	3	AC017499
14	80.5	15.0	188341	3	AC005719
15	80.5	15.0	266340	3	AE003664
16	80	14.9	6542	9	AB014524
17	80	14.9	6655	9	AY099469
18	80	14.9	12247	1	AE002239
19	80	14.9	86364	2	AC127065
20	80	14.9	151865	2	AC025197
21	80	14.9	179262	9	AL355343
22	80	14.9	184778	2	AC023567
23	80	14.9	199116	9	AP000871
24	79	14.7	8362	3	AY118524
25	78.5	14.6	106935	3	AC087079
26	78.5	14.6	263190	2	AC015657
27	78.5	14.6	303650	1	AP001519
28	78	14.5	86124	2	AC094996
29	78	14.5	110000	2	AC095071_1
30	78	14.5	123149	3	CEY66D12
31	78	14.5	208813	2	AC094998
32	77.5	14.4	134199	1	SYCSLRF
33	77	14.3	230372	2	AC073693
34	76.5	14.2	1536	5	CHKCX56
35	76.5	14.2	77555	9	AL359738
36	76	14.2	137414	2	AP004659
37	76	14.2	164153	2	AP005388
38	75.5	14.1	115604	2	AC123175
39	75.5	14.1	167692	2	AC128709
40	75.5	14.1	191832	30	AC011104
41	75.5	14.1	192961	2	AC126183
42	75.5	14.1	197143	2	AC069173
43	75.5	14.1	218836	9	AC016734
44	75	14.0	1683	6	AX250163
45	75	14.0	3813	3	AY058304

ALIGNMENTS

RESULT 1

SCCMP2
LOCUS SCCMP2 2998 bp DNA linear PLN 25-JUN-1991
DEFINITION Yeast Cmp2 gene for calmodulin-binding protein 2.
ACCESSION X54964
VERSION X54964.1 GI:3559
KEYWORDS calmodulin binding protein; Cmp2 gene.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 2998)
AUTHORS Liu, Y., Ishii, S., Tokai, M., Tsutsumi, H., Ohki, O., Akada, R.,
Tanaka, K., Tsuchiya, E., Fukui, S. and Miyakawa, T.
TITLE The Saccharomyces cerevisiae genes (Cmp1 and Cmp2) encoding
calmodulin-binding proteins homologous to the catalytic subunit of
mammalian protein phosphatase 2B
JOURNAL Mol. Gen. Genet. 227 (1), 52-59 (1991)
MEDLINE 91260679
PUBMED 1646387
REFERENCE 2 (bases 1 to 2998)
AUTHORS Miyakawa, T.
JOURNAL Submitted (30-OCT-1990) Miyakawa T., Department of Fermentation
Technology, Faculty of Engineering, Hiroshima University, Saijo,
Higashi-Hiroshima 724, Japan
COMMENT See X54963 for related sequence.
FEATURES
Source Location/Qualifiers
1..2998
/organism="Saccharomyces cerevisiae"
/strain="A364"
/db_xref="taxon:4932"
/clone_lib="lambda gt11"
539..545
/note="consensus alpha-factor response element"
663..669
/note="consensus alpha-factor response element"
807..813
/note="consensus alpha-factor response element"
907..921
/gene="Cmp2"
907..921
/gene="Cmp2"
/codon_start=1
/product="calmodulin-binding protein 2"
/protein_id="CAA38712.1"
/db_xref="GI:3560"
/db_xref="SWISS-PROT:P14747"

misc_feature
misc_feature
misc_feature
gene
CDS
misc_binding
BASE COUNT 878 a 673 c 558 g 889 t
ORIGIN
Alignment Scores:
Pred. No.: 4,54e-47 Length: 2998
Score: 537.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-09-814-661A-2 (1-104) x SCCMP2 (1-2998)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnGlnAlaPro 20
|||||
DB 110 ATGCAAAATCCCAAGACTACTTTTACCGCTCAAAATCCGCTCCACACACACACAGCCCT 169
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuProMetAla 40
|||||
DB 170 TCCACATTGCGTACCGTACCATGCGCAATTTAGAACGGTGCCTTTGCCACCTATGGCT 229
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
|||||
DB 230 GAGGTTCCTATGTCTACTCAAACTCCATGGCAGCTCCGCTTCTGCCCTCCCTTCT 289
QY 61 SerLeuGluMetTrpGluLysAspLeuGluAlaArgLeuAsnSerIleAspHisAspMet 80
|||||
DB 290 TCATTAGAAATGTGGAAAGAGATTGGAGAGACTCAACTCTATCGATCATGCATG 349
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlnLysValGlu 100
|||||
DB 350 AACACAAACAAATTTGTTCTGGCGCACTAAATCTATGTTCAACCGAGGTAAGTGCAG 409
QY 101 GluMetAspPhe 104
|||||
DB 410 GAAATGGACTTC 421

RESULT 2
SC9958 20951 bp DNA linear PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome XIII cosmid 9958.
DEFINITION 246729 271257
ACCESSION 246729.1 GI:577134
VERSION
KEYWORDS calcineurin; CNA2; CVB2; cytochrome b2; delta element; GAL80;
inosine-5'-monophosphate dehydrogenase; pif1; protein phosphatase;
transfer RNA-Tyr.
SOURCE Saccharomyces cerevisiae.
ORGANISM
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS Saccharomycetales; Saccharomycetaceae; Saccharomyces.
JOURNAL 1 (bases 1 to 20951)
AUTHORS Devlin, K. and Churcher, C.
REFERENCE 2 (bases 1 to 20951)
AUTHORS Barrell, B. and Rajandream, M.A.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-1994) Saccharomyces cerevisiae chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall Hinxton, Cambridge
CB10 1RQ by Bart Barrell and Marie-Adele Rajandream. E-mail:
barrell@sanger.ac.uk

COMMENT
All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50% of their
length by a larger CDS have been omitted from this analysis.
Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are also
included but some of these may be fortuitous. The length in
codons/amino-acids is given for each CDS as is the calculated codon
adaptation index (CAI).
Cosmid 9958 is overlapped at the start of this sequence by cosmid
9745
and at the end of this sequence by cosmid 9827.
Bases 7395 to 7399 and 14397 to 14398 are ambiguous.
FEATURES
Source Location/Qualifiers
1..20951
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 9958"
/complement(1..507)
/gene="PIF1"
/complement(<1..507)
/note="YM9958.01c, incomplete PIF1 gene, len: 169, CAI:
gene
CDS

0.14"
/codon_start=1
/product="Pif1p"
/protein_id="CAA86714.1"
/db_xref="GI:577136"
/db_xref="SWISS-PROT:P07271"
/translation="MPKWRSTLNLHPIRRPFTCSNPSNLLKNVSHAKLSFMSRRG
FRSNPITQALRKHPSILSDDDDDDEPDCIOLETEKQKKIITDTHKEDPV
DKPRMDKNVMEINKDSPLSWDMFKPSIIQPOLISENSFDQSQKKSRSRTGKFN
LRPAKKES"
846..1976
/note="Y9958.02, unknown orf, len: 376, CAI: 0.15"
/codon_start=1
/product="unknown"
/protein_id="CAA86715.1"
/db_xref="GI:577136"
/db_xref="SWISS-PROT:P53397"
/translation="MSYKFGKLAINKSELCLANVLAQGSFRWIWDEKLNQSTTWKI
GQEKYSVILRDDENEIEFVAVGCGNODALKTKMYFRLDVSLKLFNFWIP
SKFAKLSPOGIRILAQEPWETLISFICSSNNISRIITRMCNSLCNFCNLITTDG
VAYHSPTESELTSLATEAKRELGFGRYAKYIETARKLVNDKAKANITSDTYTLOS
ICKDAQYDREHLSYNGVGPVADCVCLGMLHMDGIVPDVHVSRKADRYOISAN
KNHLKRLTKYNALPISRKKNLELDHILRLMFKWGSVAGWAQGVLSKEIGGTSGS
TTTGTIKKKWDMIKETEAIVTKOMKLKVELSLDLHKEAKID"
complement(2194..7233)
/note="Y9958.03c, similar to unknown C. elegans gene
YOL4_CAEEL ZK370.4 CE00396, and E.coli YCHK SW:YCHK_ECOLI
P37053 and Shigella SW:YCHK_SHIFL P37054; len: 1679, CAI:
0.17"
/codon_start=1
/product="unknown"
/protein_id="CAA86716.1"
/db_xref="GI:577137"
/db_xref="SWISS-PROT:Q04958"
/translation="MRSNCTTNTNTGNTKNTKNSLGSFNSNNTSYRFOCTLDQI
ISAOQWSSLSLNFWSVYFVMGASRIQFRGWILAILSLRIPKWIFFKHVVQF
TLFSWILFALAVIVFYTIMKIRILSQRKLTPEFLPLENTGKSGSSANINAASTQ
SANAPAEISSTGASSIIDSXKSLDKGNETFLSYLDQFSLKAIKFGYLEKPVF
HDLTKNKTQMDGELLIDSTIGFVIVVEGTLQYHEVDSKDHGDETDRSDVFG
LDQDREDEDEEDIDNDYDKSCSNLIDEDESGVYHLKNGLGNQLLNTVKG
NPLTSLVILNLTSHMSYGVNSNFPSELSPIDITVSYNNMFCSEQNFSTDSNTN
NPSFPTSPSPKLVARATDCTIGIIPQSPAKLTAKYPRASAHIIQWLTKLHY
TFOTANDYGLTKEMDIEVLNKSIVPELYTLKAVIRKFKTVDKSGSADLEPKP
KNASAKLKKPKAPQIQLKFKGTTIVEQNARGGLFYIIISGKSVNVTNSS
ROQNFGLDLSNPLSTMDILSPNTHNNRNKSGINTSTSNKRSRSSNNAS
VHSKFSLSLSPELRNAQLSTPLSLDNTVHDIHPSPVHLKGRVSPRNLPPTGFS
AAQETEDSALRMALVEAMLYLVGNKSNMVSSSSIANNMSSLNQNMYSRRPSN
ASFMSPHCTPSDISVASSPASQPTMLRILPKYTTISNKRHNSKSDQKKPRAY
KEBLTNLDFEDVKKQFAQIQKFKGTTIVEQNARGGLFYIIISGKSVNVTNSS
SVVSSMSPQVSAQSHKGENPHHTOHLISVSGGIVGYLSSLIYGKYSFNIVAKS
DYVGFLSATLERLDKYFLYIRISDSITLKLSSRLKLHLALWHLRASSETLFS
QDGSANGIYVNLGRQLQOQSLNSNTSSEVEQNTIILGELAQGESGFGEVLTFA
MNRYSITVAVRDESLARIPTLPELLALEHPSIMIRSVLAKKIVGDRTPALTDGP
LSIKENDFTSLIPTKASVSSLSHKPQNTISCTIIFRTIITLPSGLPVEAFMKL
VQAFQVGRITIGLNTORTLHLGRHAFDRLSKOSGYFAELEEMQIVVYISDTPV
KSNWTRCQAQGDICLLADARPSAEIEYKELLSNKTARTLILHPEYVEFP
LTKHLRYRPHWSHHHIOFLTGTIMNEGKMHVNLNGALMLDKLIOTEFGRKTQQ
NISKLLPDSIKNTVENFSRPMKSKRYQYTPVHRHKNDFLRARILSGOAGLVLGG
GARGISHLGVIQAEQIPDVIGTSGSFVGLYAKDYDLVPIYGRVKFAGRIS
SIWRMLDITWPTVSTYTGHEFNRIKTFGDRIEDFWIQQYCNSTNITDSVQEIHS
FGAWRIIRASMSLAGLPLEENGMLDGGVDNLPTVMRARGCOTIFAVDVGSA
DRTPEYDGLNGFWIIFNRNPFSSHPIPNMAEIQVRLGYVASVNALEAKRNTPG
VVYVRPIIEYATLDFSKFEIYHVGVDYGRIFLQGLIDDRMPYIPGSOETTLNSQV
PEFLHRRNSI"
7395..7399
/note="ambiguous sequence, agaag or ggaag"
8358..8672
/note="Y9958.04, unknown orf, len: 104, CAI: 0.32"
/codon_start=1
/product="unknown"
/protein_id="CAA86717.1"
/db_xref="GI:577138"
/db_xref="SWISS-PROT:Q04964"

misc_feature

CDS

gene
CDS
/translation="MONSDQFYAQRNCCQQAPSTLRTVTMAEFRVPLPPMAEYPM
LSTQNSMGSSASASASLEMEKDLERLNSIDHDMNNKFGSGELKSMFNOCKVEEM
DF"
9155..10969
/gene="CNA2"
9155..10969
/note="CNA2"
/note="Y9958.05, CNA2 or CMP2 gene, protein phoshatase,
P2B2_YEAST P14747, len: 604, CAI: 0.16"
/codon_start=1
/product="calcineurin"
/protein_id="CAA86718.1"
/db_xref="GI:577139"
/db_xref="SWISS-PROT:P14747"
/translation="MSSDAIRNTEQINAAIKIENKTERPOSSTTTPIDSKASTVAAAN
STATETSRLDTQYLDGVRVSTNRRIMNKVPATISHTVPTDEELFQPGIPRIEFLRD
HFKEGKLSAAQAARIYVTLATELFSKEPNLISVPADITVCGDINGYFOLLKLFVEVGG
DPATTSILFLGDIYDRGSFCELIYLSLKNFNDHFWLLRGNHSCKHLTSTYFERN
EMLHKNLDIYKCCSFNNPLAALMNGQYLCVHGGI SPENLSQDINNLRFRIP
SHGLMCDLLWADPIEYDEVDLDEEDIVNSKTVPHHGKMAPSRDMFVPSNRGCG
SYAFTVRAACHFLOETGLLSIIRAEQAQAGYMYKNTKTLPSPSLTLTFSAPNYLDT
YNNKAAILKAYENNVMNIRQFMTPHFWLPDPDMFTWSLPFVGEKVTMLVAILNIC
TEDELENDPVIIEELVGTGKKLPQACKSEATPOPATSSAPKHASILDDERRKALRNK
ILAVAKSRMYSVLKREENTNKVQQLDHNSGVLPRGALSGVKGIDLEALSTFERARKHD
LINEKLPPSLDELKNEKNKYYEKVQKQVHEHDAKNDK"
9755..9772
/gene="CNA2"
/note="P500125 Serine/threonine specific protein
phosphatases signature"
complement(join(11169..12283,12692..13151))
/note="Y9958.06c, orf len: 524, CAI: 0.48, similar to
inosine-5'-monophoshate dehydrogenases"
/codon_start=1
/product="putative inosine-5'-monophoshate dehydrogenase"
/protein_id="CAA86719.1"
/db_xref="GI:577140"
/db_xref="SWISS-PROT:P50094"
/translation="MSAAPLDYKKALEHLKTYSKDGLSYQELMDSTTRGGLTYNDFL
VLPLVNPSSAVSLQATKTKITLNTPFVSSPMDVTADMAIYMALGGIGCFIHNN
CTPEQASAVKVKVKEFNGFPIVSPITTVGVEKVKMKRKFSGFPVTEGKCPG
KLVGLYRDIOFLDEDSLVSVMVTKNPVTGKIGITKEGNEILKQTKKGLLIYDD
NGNLVSNLSRADLMKNQNTPLASKSATTKOLLGALITGTEADKERLRLVLEAGLVV
ICTDSGNSVYQLNMKIKETFPDLEIITAGNVATREQAANLAAAGADGRLIGMSGGS
ICITQVMACGRPQGTAVYVNCVFANQFGVPCMDGQVQNIHITKALALGSSVTMMG
GLMAGTETSPGEYFYKDKRLKAYRGMSIDAMQKTGNKGNASTSRYSFSDSVLVAQ
GVSGAVVDGSGIKKFTPLYNLVQLHSCQDQIGCESLTSLENVQNGEVRFEPTASQAL
EGVHNLHSEYKELYN"
complement(11730..11768)
/note="P500487 IMP dehydrogenase/GMP reductase signature"
complement(12284..12286)
/note="splice acceptor sequence, cag"
complement(12396..12402)
/note="splice branch sequence tactaac"
complement(12686..12691)
/note="splice donor sequence gtatgt"
13765..14301
/note="Y9958.07, unknown orf, len: 178, CAI: 0.15"
/codon_start=1
/product="unknown"
/protein_id="CAA86720.1"
/db_xref="GI:577141"
/db_xref="SWISS-PROT:Q04969"
/translation="MSSAKPINVYSIPELNQALDEALPSVFARLNYERSVALLDAKLY
IGYSIAVAGLSFELKKFERDOIYVTKLLVGAYFVLSLLFWFSRFEIKGTYYVCK
RGCTKEIYVTKTEKNEPLIYVELVQKKGKSKKELKAKLEVNKVFNSGVLQDA
YKFWSEQRHVLDTKKNE"
14397..14398
/note="ambiguous sequence, ct or cc"
complement(14508..16283)
/gene="CYB2"
complement(14508..16283)
/note="CYB2"
/note="Y9958.08c, CYB2 gene, CYB2_YEAST P00175,

misc_feature

gene

CDS

cytochrome b2 precursor, len: 591, CAI: 0.21"
 /codon_start=1
 /product="cytochrome b2 precursor"
 /protein_id="CAA86721.1"
 /db_xref="GI:577142"
 /db_xref="SWISS-PROT:P00175"
 /translation="MLKYKPLKIKSKNCEAAILRASKTRTLNIRAYGSTVPKSKSFEQ
 DSRKRTQSWALRVGAILAATSSVAVLNWHNQIDNEPKLDNKKQKISPAEYAKHKKP

Alignment Scores:
 Pred. No.: 3.55e-46 Length: 20951
 Score: 537.00 Matches: 104
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-09-814-661A-2 (1-104) x SC9958 (1-20951)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 |||||
 Db 8358 ATGCAAAATTCAGACACTCTTTACGCTCAAAATCGCTGCCAACCAACAGCCCT 8417
 |||||
 QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
 |||||
 Db 8418 TCCACATTCGTACCGTACCCTGCGGGAATTTAGAAGGTCCTTTGCCACCTATGGCT 8477
 |||||
 QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerAlaSerAlaSerAlaSer 60
 |||||
 Db 8478 GAGGTTCCTATCTGTCTACTCAAACTCCATGGGCGAGCTCGCTTCCTCCGCTTCT 8537
 |||||
 QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspMet 80
 |||||
 Db 8538 TCATTAGAATGTGGAAAGGATTTGGAGGAGACTCACTATCATCATCATGACATG 8597
 |||||
 QY 81 AsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
 |||||
 Db 8598 AACACCAACAAATTTGGTCTGCGGAACATAAATCTATGTTCAACGAGGTAAGTCGAG 8657
 |||||
 QY 101 GluMetAspPhe 104
 |||||
 Db 8658 GAAATGGACTTC 8669

RESULT 3
 HSA342553
 LOCUS
 DEFINITION
 Homo sapiens genomic sequence surrounding Not1 site, clone
 NRI-063S.
 ACCESSION
 AJ342553
 VERSION
 AJ342553.1 GI:15917662
 KEYWORDS
 SOURCE
 Homo sapiens.
 ORGANISM
 Homo sapiens

REFERENCE
 AUTHORS
 Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,
 Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V.,
 Levitsky, V.G., Kolchanov, N.A., Protodopov, A.I., Kashuba, V.I.,
 Kisselev, L.L., Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.
 Not1 flanking sequences: a tool for gene discovery and verification
 of the human genome
 Nucleic Acids Res. 30 (14), 3163-3170 (2002)

JOURNAL
 MEDLINE
 22131767
 PUBMED
 12136098
 REFERENCE
 2 (bases 1 to 996)
 Zabarovsky, E.R.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
 Karolinska Institute, Tomtebodavägen 18B, S-141 86 Stockholm 171 77,
 Sweden
 FEATURES
 source
 Location/Qualifiers
 1..996

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NRI-063S"
 BASE COUNT 194 a 289 c 278 g 234 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.65 Length: 996
 Score: 86.00 Matches: 16
 Percent Similarity: 57.45% Conservativeness: 11
 Best Local Similarity: 34.04% Mismatches: 20
 Query Match: 16.01% Indels: 0
 DB: 9 Gaps: 0
 US-09-814-661A-2 (1-104) x HSA342553 (1-996)
 QY 37 ProPrometAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerAlaSer 56
 |||||
 Db 271 CCTCCAGTTCGAGGTACCAATGATACACACAGGGGTCTCTGGGGAGTTTCTCTAAC 330
 |||||
 QY 57 AlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIle 76
 |||||
 Db 331 TCTCTTCTCTGAAGACACGCTGGGAGATGAATTTCTCCAGCACCCACCTCACATG 390
 |||||
 QY 77 AspHisAspMetAsnAsn 83
 |||||
 Db 391 TACCACGGCAGGTAGAGAA 411
 |||||
 RESULT 4
 HSAJ963E22
 LOCUS
 DEFINITION
 Human DNA sequence from clone RP5-963E22 on chromosome 20 Contains
 the 3' end of a novel gene similar to NY-REN-2 Antigen, 5 Cpg
 Islands, ESTs, STSS and GSSs, complete sequence.
 ACCESSION
 AL096828
 VERSION
 AL096828.29 GI:9663381
 KEYWORDS
 HTG; Cpg Island.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 LLOYD, D.
 Direct Submission
 Submitted (05-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Aug 2, 2000 this sequence version replaced gi:9588454.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 IMPORTANT: This sequence is not the entire insert of clone
 RP5-963E22 It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP5-963E22 is at 1 in this sequence. The
 true left end of clone RP11-261N11 is at 140053 in this sequence.
 The true right end of clone RP11-305P22 is at 53294 in this
 sequence. This sequence was finished as follows unless otherwise
 noted: all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., phred
 quality >= 30); an attempt was made to resolve all sequencing

problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-963E22 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOF: pcvbac2

FEATURES

source
Location/Qualifiers
1. 163403
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="23C1-23C5"
/clone="P1s DS02190 (D82) and DS00906 (D99)"
/clone.lib="p1 library, partial Sau3A in pad10sacBII"
BASE COUNT 45206 a 35896 c 36551 g 45750 t
ORIGIN

Alignment Scores:

Pred. No.: 3.29e+03 Length: 163403
Score: 82.00 Matches: 28
Percent Similarity: 43.88% Conservative: 15
Best Local Similarity: 28.57% Mismatches: 41
Query Match: 15.27% Indels: 14
DB: 3 Gaps: 3

US-09-814-661A-2 (1-104) x AC007765 (1-163403)

QY 10 AlaGlnAsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThrMetAla 29
Db 80930 GCGAGGAGCGGAAATTCACGCGAGCAGCTCCACGGGCAATACGCCACC----- 80877
QY 30 GluPheArgArgValProLeuProMetAlaGluValProMetLeuSerThrGlnAsn 49
Db 80876 -----AAAAAGGTGCCACACCCAGCTGGTGGCGCTCCCTCCCTGAAACCGACAGTAAC 80823
QY 50 SerMetGlySerSerAlaSer-----AlaSerAlaSerSerLeuGluMetTrpGluTys 67
Db 80822 ACTGCTGGCAGTAGCAGCTCCGACGAGGACAGCTCCTCAAGTCCGCAATCCAGTTCCGAAA 80763
QY 68 AspLeuGluGluArgLeuAsnSerLeuAspHisAspMetAsn----- 82
Db 80762 TCGACGAGCTCTTCGAGCAGCAGCAGCAGCAGCAGGAAACCCAGCAACCAACCAACCGGATA 80703
QY 83 -----AspLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGly 97
Db 80702 GTCAAGCTCAACAGACTGGCGCTGTGCAGAGAGAGCGGCTCTGTGGCGACGGG 80649

RESULT 8

AC092187/c
LOCUS 164920 bp DNA linear INV 26-JUN-2001
DEFINITION Drosophila melanogaster, chromosome 2L, region 23E-23F, BAC clone
AC092187, complete sequence.
AC092187
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS
Celnikier, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Garg, N.S., George, R.A.,
Ferreira, S., Frise, E., Galle, R.F., Hoskins, R.A., Hostin, D., Howland, T.J.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclob, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 2L, region 23E-23F
Unpublished
2 (bases 1 to 164920)
Celnikier, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,

TITLE

JOURNAL
REFERENCE
AUTHORS

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Garg, N.S., George, R.A.,
Ferreira, S., Frise, E., Galle, R.F., Hoskins, R.A., Hostin, D., Howland, T.J.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclob, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Direct Submission

JOURNAL

Submitted (26-JUN-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
COMMENT
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES

Location/Qualifiers

1. 164920
/organism="Drosophila melanogaster"

/strain="y: cn bw sp"

/db_xref="taxon:7227"

/chromosome="2L"

/map="23E-23F"

/clone="BACR23N06 (D1291)"

/clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBAC3.6)"

BASE COUNT 45484 a 36619 c 37216 g 45601 t

ORIGIN

1. 164920

Alignment Scores:

Pred. No.: 3.32e+03 Length: 164920
Score: 82.00 Matches: 28
Percent Similarity: 43.88% Conservative: 15
Best Local Similarity: 28.57% Mismatches: 41
Query Match: 15.27% Indels: 14
DB: 3 Gaps: 3

US-09-814-661A-2 (1-104) x AC092187 (1-164920)

QY 10 AlaGlnAsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThrMetAla 29
Db 102746 GCGAGGAGCGGAAATTCACGCGAGCAGCTCCACGGGCAATACGCCACC----- 102693
QY 30 GluPheArgArgValProLeuProMetAlaGluValProMetLeuSerThrGlnAsn 49
Db 102692 -----AAAAAGGTGCCACACCCAGCTGGTGGCGCTCCCTCCCTGAAACCGACAGTAAC 102639
QY 50 SerMetGlySerSerAlaSer-----AlaSerAlaSerSerLeuGluMetTrpGluTys 67
Db 102638 ACTGCTGGCAGTAGCAGCTCCGACGAGCAGCAGCAGCAGCTCCTCAAGTCCGCAATCCAGTTCCGAAA 102579
QY 68 AspLeuGluGluArgLeuAsnSerLeuAspHisAspMetAsn----- 82
Db 102578 TCGACGAGCTCTTCGAGCAGCAGCAGCAGCAGGAAACCCAGCAACCAACCAACCGGATA 102519
QY 83 -----AspLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGly 97
Db 102518 GTCAAGCTCAACAGACTGGCGCTGTGCAGAGAGAGCGGCTCTGTGGCGACGGG 102465

RESULT 9

AE003581
LOCUS 306436 bp DNA linear INV 04-OCT-2000

DEFINITION Drosophila melanogaster genomic scaffold 142000013386046 section 10 of 16, complete sequence.

ACCESSION AE003581
VERSION AE003581.2
KEYWORDS GI:10727365
HTG

SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 306436)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazek,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Beriman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,L., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,R.A., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of Drosophila melanogaster

Science 287 (5461), 2185-2195 (2000)

10731132

REFERENCE 2 (bases 1 to 306436)
AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT On Oct 9, 2000 this sequence version replaced gi:7295843.

FEATURES

source

1. 306436

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="2L"

complement(<4762..>7571)

/gene="CG3119"

/map="23B3-23B3"

/db_xref="FLYBASE:FBan0003119"

/db_xref="FLYBASE:FBgn0031466"

complement(join(<4762..>7571))

7141..>7571))

/gene="CG3119"

/product="Ct10472"

CDS

/db_xref="FLYBASE:FBan0003119"
/db_xref="FLYBASE:FBgn0031466"
complement(join(<4762..>7571))
7141..>7571))
/gene="CG3119"

/note="CG3119 gene product"

/codon_start=1

/protein_id="AAF51199.1"

/db_xref="GI:7295899"

/db_xref="FLYBASE:FBan0003119"

/db_xref="FLYBASE:FBgn0031466"

/translation="MKLRSQLHLLTGFMAFVLAFLVLLVYVDVSRVTCWSTSTMT
TATTATIEDGPPRILCMVLTCPENVQSLARVYETWQRCRLIPASSEDYELGVD
GVPEFGYEDLWKNTRREGVHWEYAGDYDWFKADDDTYVVMNQHLRLRGD
NTPFEGYKMSRYNYSMGASYILSREALHREFAQAYESEVICQPKMGTEDEYM
GICMQNVGHFVDSHALDGDTPKPFMPDLLENYMSDANYTTPWLRMLSLSRVETGL
GICMSVAFHASRERMFYEFLLYHLKVPDNPQISERGHSRSLTSLDTRFPLED
NSNKDLQMSKPDNF"
<8340..>9248
/gene="CG2973"

/map="23B3-23B4"

/db_xref="FLYBASE:FBan0002973"

/db_xref="FLYBASE:FBgn0031467"

<8340..>9248

/gene="CG2973"

/product="Ct10063"

/db_xref="FLYBASE:FBan0002973"

/db_xref="FLYBASE:FBgn0031467"

8340..>9248

/gene="CG2973"

/note="CG2973 gene product"

/codon_start=1

/protein_id="AAF51198.1"

/db_xref="GI:7295898"

/db_xref="FLYBASE:FBan0002973"

/db_xref="FLYBASE:FBgn0031467"

/translation="MYLOFFALLSLALLATSOAEYNYNEKEKAANSASSSGEDFLSD
YLPSPNALNSEATPDGYVAPARQNTAGSTASVQASNLQNAASNAASVLLP
SPTPLVRHEONSEVSVSTQQQPOQVHQOQSEPLVSVLHQOQPEVFPASYSFN
YAVNDATGDIKEHSETRQGYVVRGEYSILIDEGYKRTVTYTDVHGNVAVNRVY
ALKAVVPAQVAQPTPFVARSRSVDVIRSSGAAGASENVLSGSSGSSGVS
GSDTFAEDSYANAPRLDSSGGPYA"
<10792..>12233
/gene="CG2975"

/map="23B3-23B4"

/db_xref="FLYBASE:FBan0002975"

/db_xref="FLYBASE:FBgn0031468"

join(<10792..>12233)

11952..>12233

/gene="CG2975"

/product="Ct10067"

/db_xref="FLYBASE:FBan0002975"

/db_xref="FLYBASE:FBgn0031468"

join(<10792..>12233)

11952..>12233

/gene="CG2975"

/note="CG2975 gene product"

/codon_start=1

/protein_id="AAF51197.1"

/db_xref="GI:7295897"

/db_xref="FLYBASE:FBan0002975"

/db_xref="FLYBASE:FBgn0031468"

/translation="MCGVRNKRKLEMLMLLGLAWGILLSLMKRTWQNHADRLKE
ESSPPSSORSRLRTPPTIAPSTNPPDILAAFLNETRVLCMVLTSPKTHHTRAI
HIKRWGRCNKLIEMSTKADKELGSVALNVRGYSNLWPKTAAALQYVYKHIFQYD
WELKADDDTYFIMENLRAELHAHNFREPVI FGKRFHQHKGDEMGAGYVLSKMLH
RLIKLGFSSNICTNNRYGDEVLGRGLAGVVGSDGSLRFGYFPFSLPHWP
QPPDWQPLLYTSPDNSCCSNTAISFYHNAQEFYVLEYIYKLRIFGIRNRELGP
LPKKYASSRPMSTNLOTKEENGIALIKMMPKKNVTTPNRTKTKTSKFKK"
complement(<12495..>14308)

/gene="CG18558"

/evidence="not_experimental"

/map="23B3-23B4"

/db_xref="FLYBASE:FBan0018558"

Db	119436	ACTGCTGGCAGTACAGCTCGGACGAGCACAGCTCTCAAGTGCCGAATCCAGTTTCGAA			
Qy	68	AspLeuGluAtrGLeuAsnSerIleAspHisAspMetAsnAsn-			
Db	119496	TCGAGCAGCTCTTCGACGACGAGCAGGCAGCACAGGAAACCACCAACACCACTGCCGGATA			
Qy	83	- - - - - AsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlncl y 97 : : : : : : 119556 GTCAAGCTGAACAAGATGGCGCTGTGCAGAAGAAGCGCTGCTGGCAGCGGG			
RESULT 10					
AC127287					
LOCUS					
DEFINITION	AC127287	231560 bp DNA linear HTG 14-JUL-2002			
	Mus musculus chromosome UNK clone RP23-345G7, WORKING DRAFT				
ACCESSION					
VERSION	AC127287.1	GI:21747701			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
TITLE	1 (bases 1 to 231560)				
JOURNAL	The sequence of Mus musculus clone				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 231560)				
TITLE	McPherson,J.D. and Waterston,R.H.				
JOURNAL	Direct Submission				
COMMENT	Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
	----- Genome Center -----				
	Center: Washington University Genome Sequencing Center				
	Center code: WUGSC				
	Web site:http://genome.wustl.edu/gsc/index.shtml				
	Contact: submissions@watson.wustl.edu				
	----- Project Information -----				
	Center project name: M.BA0345G07				
	----- Summary Statistics -----				
	Sequencing vector: M13; 0%				
	Chemistry: Dye-primer ET; 0% of reads				
	Assembly program: Phrap; version 0.990319				
	Consensus quality: 217459 bases at least Q40				
	Consensus quality: 222479 bases at least Q30				
	Consensus quality: 224947 bases at least Q20				
	* NOTE: This is a 'working draft' sequence. It currently				
	* consists of 30 contigs. The true order of the pieces				
	* is not known and their order in this sequence record is				
	* arbitrary. Gaps between the contigs are represented as				
	* runs of N, but the exact sizes of the gaps are unknown.				
	* This record will be updated with the finished sequence				
	* as soon as it is available and the accession number will				
	* be preserved.				
	1 1070: contig of 1070 bp in length				
	* 1071 1170: gap of unknown length				
	* 1171 2335: contig of 1165 bp in length				
	* 2336 2435: gap of unknown length				
	* 2436 3678: contig of 1243 bp in length				
	* 3679 3778: gap of unknown length				
	* 3779 4985: contig of 1207 bp in length				
	* 4986 5085: gap of unknown length				
	* 5086 6200: contig of 1115 bp in length				
	* 6201 6301: gap of unknown length				
	* 6301 7451: contig of 1150 bp in length				
	* 7451 7551: gap of unknown length				
	* 7551 8591: contig of 1041 bp in length				
	* 8592 8691: gap of unknown length				

Pred. No.:	6.39e+03	Length:	306436
Score:	82.00	Matches:	28
Percent Similarity:	43.88%	Conservative:	15
Best Local Similarity:	28.57%	Mismatches:	41
Query Match:	15.27%	Indels:	14
DB:	3	Gaps:	3

Qy	10	Ala	Gln	Asn	Arg	Cys	Gln	Gln	Gln	Ala	Pro	Ser	Thr	Leu	Arg	Thr	Val	Thr	Met	Ala	29
		:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	
D _b	119328	GGC	GAG	GAAG	CGCG	AAC	ATT	TCC	AGC	CAG	CAG	CAG	TCC	ACG	CGG	CAAT	ACG	CCCA	CC	-----	119381

	DB	-----AAAAGGTGGCCACACCCCAGCTGTCGCCGCCTCCCGCTGAACCCGACAAGTAAC	119435
	QY	50 SerMetGlySerSerAlaSer-----AlaSerAlaSerSerLeuGluMetIrpGluLys	67


```

* 8692 9829: contig of 1138 bp in length
* 9830 9929: gap of unknown length
* 11106 11106: contig of 1177 bp in length
* 11107 12308: gap of unknown length
* 12309 12408: gap of unknown length
* 12409 13578: contig of 1170 bp in length
* 13579 15213: gap of unknown length
* 15214 15313: contig of 1535 bp in length
* 15314 16620: gap of unknown length
* 16621 16720: contig of 1307 bp in length
* 16721 17912: contig of 1192 bp in length
* 17913 18012: gap of unknown length
* 18013 19490: contig of 1478 bp in length
* 19491 19590: gap of unknown length
* 19591 21253: contig of 1663 bp in length
* 21254 21354: gap of unknown length
* 21354 22616: contig of 1263 bp in length
* 22617 22717: gap of unknown length
* 22717 24535: contig of 1819 bp in length
* 24536 24635: gap of unknown length
* 24636 25951: contig of 1316 bp in length
* 25952 26051: gap of unknown length
* 26052 28591: contig of 2540 bp in length
* 28592 28692: gap of unknown length
* 28692 54605: contig of 25913 bp in length
* 54605 54705: gap of unknown length
* 54705 60263: contig of 5559 bp in length
* 60264 60364: gap of unknown length
* 60364 70420: contig of 10056 bp in length
* 70420 81037: contig of 10518 bp in length
* 81038 92725: contig of 11588 bp in length
* 92726 92826: gap of unknown length
* 92826 108137: contig of 15311 bp in length
* 108137 108237: gap of unknown length
* 108237 131589: contig of 23353 bp in length
* 131590 131689: gap of unknown length
* 131690 161250: contig of 29561 bp in length
* 161251 161351: gap of unknown length
* 161351 190403: contig of 29053 bp in length
* 190404 231560: gap of unknown length
* 231560 231560: contig of 41057 bp in length.
FEATURES
    source
    1..231560
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /chromosome="UNK"
        /clone="RP23-345G7"
    misc_feature
    1..1070
        /note="assembly_name:Contig18"
    misc_feature
    1171..2335
        /note="assembly_name:Contig20"
    misc_feature
    2436..3678
        /note="assembly_name:Contig23"
    misc_feature
    3779..4985
        /note="assembly_name:Contig25"
    misc_feature
    5086..6200
        /note="assembly_name:Contig28"
    misc_feature
    6301..7450
        /note="assembly_name:Contig29"
    misc_feature
    7551..8591
        /note="assembly_name:Contig31"
    misc_feature
    8592..9829
        /note="assembly_name:Contig34"
    misc_feature
    9930..11106
        /note="assembly_name:Contig38"
    misc_feature
    11207..12308
        /note="assembly_name:Contig42"
    misc_feature
    12409..13578
        /note="assembly_name:Contig47"

```

```

misc_feature
13679..15213
/note="assembly_name:Contig49"
misc_feature
15314..16620
/note="assembly_name:Contig51"
misc_feature
16721..17912
/note="assembly_name:Contig53"
misc_feature
18013..19490
/note="assembly_name:Contig54"
misc_feature
19591..21253
/note="assembly_name:Contig55"
misc_feature
21354..22616
/note="assembly_name:Contig56"
misc_feature
22717..24535
/note="assembly_name:Contig58"
misc_feature
24636..25951
/note="assembly_name:Contig59"
misc_feature
26052..28591
/note="assembly_name:Contig60
clone_end:SP6
vector_side:right"
misc_feature
28692..54604
/note="assembly_name:Contig68
clone_end:T7
vector_side:left"
misc_feature
54705..60263
/note="assembly_name:Contig61"
misc_feature
60364..70419
/note="assembly_name:Contig62"
misc_feature
70520..81037
/note="assembly_name:Contig63"
misc_feature
81138..92725
/note="assembly_name:Contig64"
misc_feature
92826..108136
/note="assembly_name:Contig65"
misc_feature
108237..131589
/note="assembly_name:Contig66"
misc_feature
131690..161250
/note="assembly_name:Contig67"
misc_feature
161351..190403
/note="assembly_name:Contig69"
misc_feature
190504..231560
/note="assembly_name:Contig70"
BASE COUNT 62325 a 52521 c 51715 g 62079 t 2920 others
ORIGIN
Alignment Scores:
Pred. No.: 6.06e+03 Length: 231560
Score: 81.00 Matches: 30
Percent Similarity: 43.48% Conservative: 10
Best Local Similarity: 32.61% Mismatches: 44
Query Match: 15.08% Indels: 8
DB: 2 Gaps: 4
US-09-814-661A-2 (1-104) x AC127287 (1-231560)
QY 8 PheTyrAlaGlnAsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThr 27
Db 147736 TTCCTGAATTCATAACATCTGCTCCTCAATGCTCAGGCGCTTCACCACTCAGGACCTTTGGA 147795
QY 28 MetAlaGluPheArgValProLeuProProMetAlaGluValProMetLeuSerThr 47
Db 147796 ATGAACATCTAT-----CCTCTCGCTCCTCAGGCAAGTGTGGTGTGCTTACGTTAT 147846
QY 48 GlnAsnSerMetGlySerSerAlaSerAlaSerAla-----SerSerLeuGluMetTrp 65
Db 147847 GTGAGTGGGTGGTCATTCAGTATGACTAGACCCCTTCCTCCCTTCCTGTCATTATGG 147906
QY 66 GluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMetAsnAsnLysPhe 85
Db 147907 AAGAAGGGCTTGCATCCCAAGG-----GCTGTTGATAGA---CTTACAATGGAACACTG 147957
QY 86 GlySerGlyLeuLeuLysSerMetPheAsnGlnGly 97

```

```

Db      147958 TCAAAATGCTCCTCTTTGTGCATAATAGGCAAGG 147993
RESULT 11
AC079431
LOCUS   AC079431          300695 bp    DNA    linear    HTG 01-SEP-2000
DEFINITION Mus musculus chromosome 16 clone RP23-1J12, WORKING DRAFT SEQUENCE,
77 unordered pieces.
ACCESSION AC079431
VERSION   AC079431.1 GI:9958043
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    Mus musculus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS   DOE Joint Genome Institute.
TITLE      Sequencing of Mouse
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 300695)
AUTHORS   DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL    Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-23_1J12
-----
Summary Statistics
Consensus quality: 164936 bases at least Q40
Consensus quality: 197223 bases at least Q30
Consensus quality: 214345 bases at least Q20
Estimated insert size: 257300; agarose-fp estimation
Estimated insert size: 293095; sum-of-contigs estimation
Quality coverage: 3.01 in Q20 bases; agarose-fp estimation
Quality coverage: 2.65 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1077: contig of 1076 bp in length
* 1177: gap of unknown length
* 2209: contig of 1032 bp in length
* 2309: gap of unknown length
* 3528: contig of 1219 bp in length
* 3628: gap of unknown length
* 4765: contig of 1137 bp in length
* 4865: gap of unknown length
* 5953: contig of 1088 bp in length
* 6053: gap of unknown length
* 7183: contig of 1130 bp in length
* 7283: gap of unknown length
* 8421: contig of 1138 bp in length
* 8521: gap of unknown length
* 9617: contig of 1096 bp in length
* 9717: gap of unknown length
* 11130: contig of 1413 bp in length
* 11230: gap of unknown length
* 12408: contig of 1179 bp in length
* 12509: gap of unknown length
* 13768: contig of 1259 bp in length
* 13868: gap of unknown length
* 15044: contig of 1177 bp in length
* 15144: gap of unknown length
* 16325: contig of 1181 bp in length
*
* 16326: gap of unknown length
* 16426: contig of 1153 bp in length
* 17578: gap of unknown length
* 17679: contig of 1167 bp in length
* 18945: gap of unknown length
* 20060: contig of 1115 bp in length
* 20160: gap of unknown length
* 21258: contig of 1098 bp in length
* 21358: gap of unknown length
* 21359: contig of 1032 bp in length
* 22490: gap of unknown length
* 22491: contig of 1109 bp in length
* 23600: gap of unknown length
* 23699: contig of 1208 bp in length
* 24907: gap of unknown length
* 24908: contig of 1114 bp in length
* 26121: gap of unknown length
* 26222: contig of 1052 bp in length
* 27274: gap of unknown length
* 27374: contig of 1099 bp in length
* 28472: gap of unknown length
* 28473: contig of 1111 bp in length
* 28573: gap of unknown length
* 29684: contig of 1051 bp in length
* 30835: gap of unknown length
* 30935: contig of 1101 bp in length
* 32035: gap of unknown length
* 32135: contig of 1199 bp in length
* 33334: gap of unknown length
* 33435: contig of 1184 bp in length
* 34598: gap of unknown length
* 34699: contig of 1130 bp in length
* 35818: gap of unknown length
* 35819: contig of 1164 bp in length
* 37082: gap of unknown length
* 37183: contig of 1463 bp in length
* 38646: gap of unknown length
* 39768: contig of 1023 bp in length
* 39769: gap of unknown length
* 39869: contig of 1105 bp in length
* 40974: gap of unknown length
* 41073: contig of 1101 bp in length
* 42174: gap of unknown length
* 42274: contig of 1087 bp in length
* 43341: gap of unknown length
* 43442: contig of 1051 bp in length
* 44533: gap of unknown length
* 44633: contig of 1436 bp in length
* 46069: gap of unknown length
* 46169: contig of 1253 bp in length
* 47421: gap of unknown length
* 47521: contig of 1108 bp in length
* 48629: gap of unknown length
* 48729: contig of 1257 bp in length
* 49986: gap of unknown length
* 50087: contig of 1308 bp in length
* 51395: gap of unknown length
* 51495: contig of 1664 bp in length
* 53159: gap of unknown length
* 54802: contig of 1544 bp in length
* 54803: gap of unknown length
* 54903: contig of 1107 bp in length
* 56109: gap of unknown length
* 56110: contig of 1341 bp in length
* 57453: gap of unknown length
* 57553: contig of 1113 bp in length
* 58668: gap of unknown length
* 58768: contig of 1157 bp in length
* 59925: gap of unknown length
* 60025: contig of 1135 bp in length
* 61160: gap of unknown length
* 62660: contig of 1505 bp in length
* 62765: gap of unknown length

```


This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
http://www.wormbase.org/db/seq/sequence?name=T05C3;class=Sequence

NEIGHBORING COSMID INFORMATION

The 5' cosmid is ZC317, 200 bp overlap; the 3' cosmid is C24C6, 200 bp overlap. Actual start of this cosmid is at base position 197 of T05C3; actual end is at 33089 of T05C3.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.dbbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (http://wormdb.dfci.harvard.edu/), similarity to other proteins from Blastx analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucleic Acids. Res., 25, 935-964).

FEATURES

source

1. 33089

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="v"

/clone="T05C3"

368. .4353

/gene="T05C3.3"

/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T05C3.3;class=Sequence"

join(368. .1975,2702. .3255,3720. .4353)

/gene="T05C3.3"

/codon_start=1

/product="Hypothetical protein T05C3.3"

/protein_id="AAB65360.1"

/db_xref="GI:2291242"

/translation="MANKEINTWKSARKAKAILAWNDVVVKEACADYGRKLENFLMEW
EETQDAKLPFGNPLDEETKLYITNIEHVMYKGEKLEELROTQKIOLEHK
VETWENMDMTSGNDGSLTQIDTPMPTARLNQNSPEDKTNENDDOETMLINTEPK
TSDPTTASYUROMQODTERKQVIAEYKQYRMOKKLDGLEOKLDPKNAKSDVQNA
KLSLELQVMOQITKNQSSSTAQENKLTSTRPGNAGTMRASAEQDGGTFGFD
EYFQGSPEKYNNFVIGVESQIHNHPREFSLQKQMAITQLLVGAKQYITNGHSHENY
YLITNLTFLVFOEFQKDYLTRKRMNTFKNSTIEVEESLATYAGDVRRLLKGLG
YVDYKYLQETDALPKIRKRYEAYNNSGNTRELLITCTQAVIAOLKAQENREME
EKVVEKSVKPARVKNVSAIRYNNNDVIMAVRDLOKYNLSLKEITRLRQKKIR
EEREEDVLTQRTAATENCQAGEFIAQIKSCDELOHNRRLQREKERIWFREE
YKLLQKQTEYEEQKMKIEVEVQDKNNIHNLKARNSHFAEVRLEHLEULR
EQEAREQQEHADQVEDQEMVAEKIIPFTCHVHKRSYIAALRDVLRNKKATSR
KLEIGQVLLVFOELPRYAMPMAIRVNLNKNPDGEVQSVYMKGRIRIVEKPVNLLI
PLENTGRIQEAQEEVPGDVMAEDVDADEQAGTAENPEEDIPKEGAVAEQDQVD
VDKQPEERENGRAHLKRAAKTRPLIRPAKFRFSPFGVSRRLPREGKEEIDNRVQ
RSAPASTEMQNVRFEK"

complement(7239. .15093)

/gene="T05C3.4"

/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T05C3.4;class=Sequence"

complement(join(7239. .7276,7318. .7506,7777. .8020,

9617. .9798,9873. .10116,10883. .11098,12131. .12247,
12294. .12505,12784. .12966,13040. .13196,13324. .13563,
13613. .13701,13759. .13879,14153. .14197,14258. .14411,
14985. .15093))
/gene="T05C3.4"
/note="contains weak similarity to Saccharomyces
cerevisiae RNA binding protein (GB:D37935)"
/codon_start=1
/product="Hypothetical protein T05C3.4"
/protein_id="AAB65362.1"
/db_xref="GI:2291244"
/translation="MNCQCNQSCSPPAQLGCLCPMPQYQQPTNYQTPSPSVIYP
SYQNOQLQPPSSSYLANIPPEMOSYPOPLFTTPPLPDNSILWIPDQIKRLIYK
INYPPTQTEIPQOTLKTSYSTSGSLGPPAPSAEKLNOVPPALKOKEEESDQOVNE
KNYVGAHQPLSDIFERYGSEADNEIDQNSYPIITTEAVYRPIETIAKRYETSTH
RTSSSRATKQESRRYETELMRGEGSDISSGYGFSASSTSPILIELEQPTKYE
NRSNTELDADVTAFTQEVDEFYKSTSPQAQTEYGAVRDARDSTGKDETSTPKFF
DRDNVNTKSRRIKREKINSCNNPILKDLMMKMTTSPSKQMYSAAYEMMM
GRNVNCSKHSFVYVNLVYITNTLLSRLLELPCFIFLFTFKSVPGFLTWTYF
AWFLHIOFLSAASLTVHRISAIYMPYKDFKSYLLHCGLAGALYSFLPTFFWLD
AIKVEIQNETLTVIYPEVLAKANNVTAITVLYFIPIFIALGVNASYFQKQILST
MHEVSKKLRIATYCIYFTGTSWLTSLNNLFLPLDFLRFIQLLLYSDDL
WSKSTIFGLAFALYSFLPTTIFGFTNEVHILNLTMRNRPDLVKTATNTAIFTI
IYFIICVIGIKTSVIINKKQASAMYEKIAKLTHTAIIVHCVVFAGILLASVLTSL
NSVNFDFDIIGINQTLIMVFSDDLGHILSASRSIF"
17335. .19955
/gene="T05C3.6"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T05C3.6;class=Sequence"
join(17335. .17360,17408. .17497,17545. .17705,18173. .18333,
18379. .18676,18725. .19037,19080. .19208,19759. .19955)
/gene="T05C3.6"
/note="coded for by the following C. elegans cDNAs:
yk723a1.3, yk723a1.5, yk735c4.3, yk735c4.5"
/codon_start=1
/product="Hypothetical protein T05C3.6"
/protein_id="AAK71397.1"
/db_xref="GI:14625255"
/translation="MPKCAETCEVVCESIPAGMTFNSSYPINFSTTDCWRLMSEAK
EELIGSYVSLLVKDGDTTDTNTWNVGYSIDLKPSANLEDLGKGAQIYNL
LSTAVNRNIVRIAGTYEAGGYPTDGLVKNNGRIVRSLDFRNWYIGGLITKSKNA
IDGKHFIYSANFDELSTNVNKELVGAVFPCPLANDLNLEIYWNKAPGATIPKO
WNSLSTPANQHSQSPMSVFOPTGQSAQYISAPGCGACGREDDTSMIKTIDEAQEYL
YMAVDYSPSTLYLKNANKPKPEONAIRRAAFRAVRVFMVSLMDITYSDVYGLLY
SLODISDILCYKWNDSNDKCLKKSIERFVQVPEMOKGKIPARYHYNKYVFTESRA
YIGTSNNSDDYQWITAGIVIRADDTSKVQQTFSVFERDWSSTYIPLKNTFI
SGNRTSNFY"
20112. .29432
/gene="T05C3.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T05C3.2;class=Sequence"
join(20112. .20330,20375. .20472,20517. .20619,20837. .21157,
21208. .21495,21667. .22188,22376. .22681,22725. .22962,
23019. .23426,23472. .23644,23693. .23823,24085. .24174,
24223. .24410,24458. .24642,24903. .25016,25066. .25168,
25569. .25676,25723. .25916,26213. .26398,26444. .26715,
26825. .27083,27136. .27303,27854. .27997,28050. .28167,
29110. .29371,29423. .29432)
/gene="T05C3.2"
/note="contains similarity to Vaccinia virus 37 kd
envelope protein"
/codon_start=1
/product="Hypothetical protein T05C3.2"
/protein_id="AAK71396.1"
/db_xref="GI:14625254"
/translation="MYSSISRSVHKKRSKSTSTSQLTRSSKSSKSTESKFLIPNHIQ
KKNGNSRTTGLDMLLSQNNQKQVVERRDYQTEIFDRVESCGLNIPKPSKLVK
VFTSSTPTDTTVERNALMEELYPKKEYCRDYGLDFQVDMRGVDESDHMTTK
LCILSEIANCORLSVGANFVFLCOYGYRIPSEILSTELMLKRLTREHEDISLLD
NWTIEDSNSVPPMFVLOQISSILVFNKKRIPKIQEQADAKQWCECAKMOOLLKKGAR

/gene="CG10447"
/map="38A2-38A2"
/db_xref="FLYBASE:FBan0010447"
/db_xref="FLYBASE:FBgn0032816"
join(<4094, 4166, 4231, >4709)
/gene="CG10447"
/product="CT29338"
/db_xref="FLYBASE:FBan0010447"
/db_xref="FLYBASE:FBgn0032816"
join(<4094, 4166, 4231, >4709)
/gene="CG10447"
/note="CG10447 gene product"
/codon_start=1
/protein_id="AAF53839.1"
/db_xref="GI:7298623"
/db_xref="FLYBASE:FBan0010447"
/db_xref="FLYBASE:FBgn0032816"
/translation="MNSSEDSQQYLNDMLVKEDEASGDDSKQDGGIMLREQDRFLP
ICNIIKIMKVPVQNPQKIAKARECIEOECVSEFISFISSEIAERVAENRKTNGDDL
LVASNLGDFNYPEPLSIYLQRYREVSEITLCTAVNNQVTHCSRTRNIVISFWTPVIR
TTKMGPPQTMRGNSNPLTYPVs"
complement(<11183, >24830)
/gene="CG10631"
/map="38A2-38A3"
/db_xref="FLYBASE:FBan0010631"
/db_xref="FLYBASE:FBgn0032817"
complement(join(11183, 11607, 11666, 17621, 17682, 20176,
20236, 21318, 21392, 21650, 21710, 22039, 22098, 22243,
23832, >24830))
/gene="CG10631"
/product="CT29778"
/db_xref="FLYBASE:FBan0010631"
/db_xref="FLYBASE:FBgn0032817"
complement(join(11183, 11607, 11666, 17621, 17682, 20176,
20236, 21318, 21392, 21650, 21710, 22039, 22098, 22243,
23832, >24830))
/gene="CG10631"
/note="CG10631 gene product"
/codon_start=1
/protein_id="AAF53840.1"
/db_xref="GI:7298624"
/db_xref="FLYBASE:FBan0010631"
/db_xref="FLYBASE:FBgn0032817"
/translation="MSQOPHHAHHYAHYPPVPTPMSLQOQQOQQOQQOQLSPQ
AQOQHANWYSHVASYPTPHSAFCAPAPASCKAANNSSGNNNNMTGQGGYGGGG
AQQYGAAGGLVNSGAVGGGPNYGLGANTVAYAHNOLLOQQOQQOQQOQQOQQO
QOHOHLPOHIQOQPYMGHNTMGSPYIKSEPEAYQOPNPAPAPAPPEVLKS
EPIDERSKSNYIDNTPTADFSEFSDMLSPKVELTVKDESYTSSFLRRKQ
QDRNESLPCORCKEVFFKQVYLHVAESNCGIOEYDFKCTCPMSFMTTEELQR
HKLHRADRFCHYCGKHFDIAECEAHEYMGEYDSFVNCMSGTFFATREOLYHL
POHFOORPCICRLWYQTALEHLEHRLAAPYFCGKY YTGOSSSASQSOAQOHTN
YKLODCHMATMEPTTPHHKTTSGSLPATAALNSLLOQOANADGAAMFAASAMKN
EVNWKERSYNSSTSESYSVQSYNNAYGSSSHAGALGPQAHSTLDDSEDL
CCVPLGVRKASTPTLOFTTFPKDEKTLNQLNHLNLMKFHIPAASYANFRICSWHFPR
CINRYSUCYWAFTNLDHDDVANLYONRELTFTTGEVARCSMPHCTSORGESNLK
FYNPXDKSLIKWNCNARLPVQAKPEHFCFSRHFECIGKFLKPWAVPTLHLQAY
YRKIHONKLYVEERKCLNCRSSDFNMSLYRFPDEVLRLRWYCNLRLDQV
YRGNHKKICSAHFKAIGALGKLSPGAVPTLHLGHTDTFNIYENELWPPPTAPNSHS
GLOHQTQOHSOHLQOOLHSYHRQSAASTSSASSANSYHVPDLASLYGLGNS
GSSAMASDSNCCVPCSPCRKHNNENITFTIPRPEQMRKWNKLNK1PEKMHKG
MRICSLHFEPYCIGGCCPFAPVPTNLGHDHDDIHRPNPDVIRKLNTRCCVACVRN
RDRHANLFRFNSGNSLLTWCNQLQRPVDPGSKLFENDACEVHFEERCLNRKLEKW
AVPTLSLGHENIPIPTPQVTFEYSRPTAPNNGEBOGECVETCKRNPSSVDIKLY
RPEEASLAKWAHNLQTESOLTSMRCLNLFHFAHCIGKMRPWPATPTNLGHTTEN
LYENPESHMLYKRTIMKAKQASVKTWVPRCLPCHRCVRAHNVQLYRFPKLNR
TLKWAHNLQVPMVGSQRCLCSAHFEPHLSKKNCPVPLAVPTLDUNAPGLKIYNP
ALKASKLCIORCVI VESCRKTRAQVQLFRLPHSPQLRKMWHNITRPRAAWRAQY
RVCSEHFEHSFNGRLSAGAIPTLELGHDEDIYPNEAQAFADHCHVEGCEASKEQ
PDLVLFPTPEDDDMLWKCNLKNMPDCTGVCRICKHFEADCIGPKHLYKWAIPTE
ELGHDADIELVLPKPEDRYVDFKCIPTCGKTRFDEVOVMSFPKDALNLFEPKH
HNLRLHLSHERDRYKICNSHEDICIGKTRNLIGSIPLELGHDETDLDLVQNPAAE
LQSNLFRGRHDESGGIIILIKFESESEVDKTDVSDAKDFNTRVQKLRLTMSDLKCC
VRSGRSRLHGARLFPPTGKQOHLKWRHNLRLHDEVDVDRSTRICSAHFNRRICDK

QLRSWAMPTQOLGHEQIYENPKNIPFFPTPTCALSHCKRRRSIDNDRTRYPRSE
DLLEKWRANRLAPDCRGRICADHFESQVRGLKLTGAVPTNLKGDGLIYDNEA
IKVGMTTEEGSSSEMTRLKPKRELIDEEELDAEERPHEDHMDYDEDEKGHYFDPLE
LVETFAHPSDDEAEYRGDEEDERDEDEDIDEAEHFPDNPPTPTPLRREKANNV
TPICCLRHCKERTAFHLLSTFGPKDRQLLWLANHLNPDCCIGRVCIEHFQPEV
LGRKLRQNAVPTLVGHDEPLRYSCNMDLQDQLOQHSVFLWLSLKHKRKRKLFE
PPDIRONKSAEVRKMRQRLRMKMKIGREIKLEVOTEREMKSKETPTTSLGRIKLE
ICGINSICANDVNLQLOLPEDQNLRLKWNILKLEVDTFKEIQVCLKHFEQVQNG
KPLEOAVPTLQDQNSWNIYRNSGNCILFPCSNSSSDQLSFVLDLPGNAVLRDWSHL
NLPPSSSEGLLCSHEMOLFQVEYPKVLAQDLDLQIVDELRCAPVSCSKSDGDL
QLIPFPKDATLLKWLONTKISVDHLKHKSYRVCVLFHETCLEANFPKAWIPTLHL
NHDELNLNRPESRGTSPNSRLTPLRITDITLSGSPSCSPSPSPGRIRICCTIS
TCQIGSSQVRLYRFPTEQALLRWLVNTQOQPRIVPAELYVCSHFEPDAICKQL
RCWAEPTNLGHGDFVIPNAKHNGNIAGGDTHEAMRLIRERYCSVLTCFQAAASGVR
LYEVPKDMPTIRKAAACRHSQMASSGHFKVQSFHAFCECFEDPTNLNLEGSVPTLE
LSRGDLERHCLVSCDEKASEGLRYVYKPTTAQLNANSNLKIQQDLGLGQLLIC
ERFPEFCFAHGLRPGALPTLMGHDEEVEMLPNEPILWQKAEVCCACGRIMQ
PGPKPSGFPKTLALAKKLIHNLQSVNEDOLGSLKVCYSVHFASLFGSLTSGSIP
TLEIGHSSQDIFQSDIQNMGSLKFERRAIVSDLDICYPECKELCKTVSNLQPEEDL
RTAWLRHLKIKEPSSGVQCLPLHYVILYELSIKSFPEHVSNRLEENYLSARNRRV
KIYSCAVKGCMIIRPRDKVLLHGLPQRKIDILRMVVENGOLEITPEQOQYMLRVCNHF
ESRCSDDLRLHPSVPTLNLPGNVHQIPTKEWQEMTILKTEQADATIKLEREEELE
EQEPLEEEAGDCSLLEPIVRMEHTSEDESEMALEVLLEVGHVERMDSVDESY
TEQAVYOSNGVRNOYNNANCAVEGCVTVEDVGTIKLHKFPASSEAAARWMMHTQVD
MDEKFWMYRIRYCSHFEEQCFQSARIKKGAMPTLLGPKRKPQYVDFALQETFEELI
LPEDLEPEDTKPKREVIKCLPTPAPRKSSKPCQTEGCMNHLTTENITLHKRPHSE
DMCLKWQHNQVPPDPFHRWRYRCSVHFHPVCLLNMRLVHGSVTLKLGKSPTELF
DNDFEALNLRDLKLGTDHSTVQIKKEDEDSMPSLEPEPQLHEDQVEDESEAMQIPLN

Alignment Scores:
Pred. No.: 7.93e+03 Length: 266340
Score: 80.50 Matches: 27
Percent Similarity: 46.51% Conservative: 13
Best Local Similarity: 31.40% Mismatches: 33
Query Match: 14.99% Indels: 13
DB: 3 Gaps: 3

US-09-814-661A-2 (1-104) x AE003664 (1-266340)

Qy 20 ProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro----- 37
Db 99460 CCGGATGGCTTACGACCTCTCTCTCGCGGAATCAAAACACCACCTCCGCCCAAGG 99401
Qy 38 -----ProMetAlaGluValProMetLeuSerThrGlnAsnSerMet 51
Db 99400 GGTAAACGATTTTGGGTGCAATTTTAGAGCCACCCCTCGCAGT----- 99353
Qy 52 GlySerSerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGlu 71
Db 99352 GCATCCACGATCCGATCCGATCAGCCTCGCATCGCATCGCGCAAGGATACGCAAGCA 99293
Qy 72 ArgLeuAsnSerIleAspHisAspMetAsnAsnAsnLysPheGlySerGlyLeuLys 91
Db 99292 TGCCTG---ACAGTTGACAGCTTAGCGCTCGCGCAACAAGTTCTCTCGGAAAGAGTGGCGG 99236
Qy 92 SerMetPheAsnGlnGly 97
Db 99235 AAAATGCGCAAGAGGG 99218

Search completed: July 17, 2003, 18:39:58
Job time : 1537 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 12:00:34 ; Search time 63 Seconds
(without alignments)
4365.835 Million cell updates/sec

Title: us-09-814-661A-1

Perfect score: 2069

Sequence: 1 aatgagcaacogtgaaca.....cctcttgatcaatgcttt 1158

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 903798

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09814661/runat_16072003.115302.14701/app_query.fasta_1.1351
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.coi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09814661@cgn_1.1.17 @runat_16072003.115302.14701
-NCPU=6 -ICPU=3 -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	537	26.0	104	10	US-09-814-661A-2 Sequence 2, Appli
2	90	4.3	525	9	US-10-028-072-56 Sequence 56, Appl
3	90	4.3	525	9	US-10-121-049-56 Sequence 56, Appl
4	90	4.3	525	9	US-10-123-904-56 Sequence 56, Appl

5	90	4.3	525	9	US-10-140-470-56 Sequence 56, Appl
6	90	4.3	525	9	US-10-175-746-56 Sequence 56, Appl
7	90	4.3	525	9	US-10-176-918-56 Sequence 56, Appl
8	90	4.3	525	9	US-10-176-921-56 Sequence 56, Appl
9	90	4.3	525	9	US-10-137-865-56 Sequence 56, Appl
10	90	4.3	525	9	US-10-140-474-56 Sequence 56, Appl
11	90	4.3	525	9	US-10-142-431-56 Sequence 56, Appl
12	90	4.3	525	9	US-10-143-114-56 Sequence 56, Appl
13	90	4.3	525	9	US-10-140-002-56 Sequence 56, Appl
14	90	4.3	525	9	US-10-142-419-56 Sequence 56, Appl
15	90	4.3	525	9	US-10-123-262-56 Sequence 56, Appl
16	90	4.3	525	9	US-10-142-423-56 Sequence 56, Appl
17	90	4.3	525	9	US-10-121-050-56 Sequence 56, Appl
18	90	4.3	525	9	US-10-141-755-56 Sequence 56, Appl
19	90	4.3	525	9	US-10-143-032-56 Sequence 56, Appl
20	90	4.3	525	9	US-10-123-108-56 Sequence 56, Appl
21	90	4.3	525	9	US-10-123-236-56 Sequence 56, Appl
22	90	4.3	525	9	US-10-123-261-56 Sequence 56, Appl
23	90	4.3	525	9	US-10-140-921-56 Sequence 56, Appl
24	90	4.3	525	9	US-10-140-928-56 Sequence 56, Appl
25	90	4.3	525	9	US-10-121-045-56 Sequence 56, Appl
26	90	4.3	525	9	US-10-123-292-56 Sequence 56, Appl
27	90	4.3	525	9	US-10-123-903-56 Sequence 56, Appl
28	90	4.3	525	9	US-10-124-819-56 Sequence 56, Appl
29	90	4.3	525	9	US-10-124-822-56 Sequence 56, Appl
30	90	4.3	525	9	US-10-140-925-56 Sequence 56, Appl
31	90	4.3	525	9	US-10-160-498-56 Sequence 56, Appl
32	90	4.3	525	9	US-10-121-041-56 Sequence 56, Appl
33	90	4.3	525	9	US-10-121-043-56 Sequence 56, Appl
34	90	4.3	525	9	US-10-121-047-56 Sequence 56, Appl
35	90	4.3	525	9	US-10-123-215-56 Sequence 56, Appl
36	90	4.3	525	9	US-10-123-902-56 Sequence 56, Appl
37	90	4.3	525	9	US-10-123-908-56 Sequence 56, Appl
38	90	4.3	525	9	US-10-123-909-56 Sequence 56, Appl
39	90	4.3	525	9	US-10-123-910-56 Sequence 56, Appl
40	90	4.3	525	9	US-10-124-813-56 Sequence 56, Appl
41	90	4.3	525	9	US-10-124-817-56 Sequence 56, Appl
42	90	4.3	525	9	US-10-124-824-56 Sequence 56, Appl
43	90	4.3	525	9	US-10-125-922-56 Sequence 56, Appl
44	90	4.3	525	9	US-10-125-924-56 Sequence 56, Appl
45	90	4.3	525	9	US-10-127-825A-56 Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-814-661A-2
; Sequence 2, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814, 661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 104
; TYPE: PRT
; ORGANISM: S. Cerevisiae
US-09-814-661A-2

Alignment Scores:
Pred. No.: 8.66e-52 Length: 104
Score: 537.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.95% Indels: 0
DB: 10 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-814-661A-2 (1-104)

QY 549 ATGCAAAATTCACAGACTACTTTTACGCTCAAAATCGCTGCCAACACAAACCCCT 608
 Db 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnA9CySgInGInGlnAlaPro 20
 QY 609 TCCACATTGCGTACCGTGACCATGCGGGAATTTAGAGGGTCCCTTTGCCACTATGGCT 668
 Db 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPrometAla 40
 QY 669 GAGGTTCTATGTTCTACTACTCAAACTCCATGGCAGCTCGCTTCCCTCCCTTCT 728
 Db 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
 QY 729 TCATTAGAAATGTGGAAAGGATTGGAGGAGAGACTCAACTCTATGCATCATCACATG 788
 Db 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
 QY 789 AACAAACAAATTTGGTCTGCGCAACTTAAATCTATGTTCAACCGAGGTAAGTCCGAG 848
 Db 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
 QY 849 GAAATGGACTTC 860
 Db 101 GluMetAspPhe 104

RESULT 2

US-10-028-072-56

Sequence 56, Application US/10028072

Publication No. US20030004311A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
 APPLICANT: Beresini, Maureen
 APPLICANT: DeForge, Laura
 APPLICANT: Desnoyers, Luc
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Sherwood, Steven
 APPLICANT: Smith, Victoria
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Watanabe, Colin K
 APPLICANT: Wood, William
 APPLICANT: Zhang

TITLE OF INVENTION:

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/028,072

CURRENT FILING DATE: 2001-12-19

PRIOR APPLICATION NUMBER: 60/049911

PRIOR FILING DATE: 1997-06-18

PRIOR APPLICATION NUMBER: 60/056974

PRIOR FILING DATE: 1997-08-26

PRIOR APPLICATION NUMBER: 60/059113

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059115

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059117

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059122

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059184

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059352

PRIOR FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/059588

PRIOR FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/059836

PRIOR FILING DATE: 1997-09-24
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/062285
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/062287
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/062814
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/062816
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063045
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063082
 PRIOR FILING DATE: 1997-10-31
 PRIOR APPLICATION NUMBER: 60/063127
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063327
 PRIOR FILING DATE: 1997-10-27
 PRIOR APPLICATION NUMBER: 60/063329
 PRIOR FILING DATE: 1997-10-27
 PRIOR APPLICATION NUMBER: 60/063550
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063561
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063704
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063733
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063735
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063738
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063755
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064248
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/064809
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: 60/065186
 PRIOR FILING DATE: 1997-11-12
 PRIOR APPLICATION NUMBER: 60/065846
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/066453
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/066511
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/066770
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/069212
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069278
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069334
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069694
 PRIOR FILING DATE: 1997-12-16
 PRIOR APPLICATION NUMBER: 60/072320
 PRIOR FILING DATE: 1998-01-23
 PRIOR APPLICATION NUMBER: 60/073612
 PRIOR FILING DATE: 1998-02-04
 PRIOR APPLICATION NUMBER: 60/074086
 PRIOR FILING DATE: 1998-02-09
 PRIOR APPLICATION NUMBER: 60/074092
 PRIOR FILING DATE: 1998-02-09
 PRIOR APPLICATION NUMBER: 60/077791
 PRIOR FILING DATE: 1998-03-12
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/079294
 PRIOR FILING DATE: 1998-03-25

;
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081695
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081818
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082999
; PRIOR FILING DATE: 1998-04-24
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085149
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086414
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086430
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088730
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088741
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538

;
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

Alignment Scores:
Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661A-1 (1-1158) x US-10-028-072-56 (1-525)

QY 32 ACCGGCTACTTATTCGCC-----AAGGATCAGCTT 61
Db 16 SerGlyPheLeuTyPProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CCTTCTGCCAACATCATTGCCGTCGAAACGTCGCGGCTCTTTCTGACATGGT----- 115
Db 36 ProAsnPheValIleIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGANTACTTCCAACTAAGAG----- 136
Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTTCTCTTTTTCGTAGCCCAATGATAGAAAG-----AACAAAT 181
Db 75 ValAspPheHisAlaAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAAATACGTACAGATATAGTAGATATGTTTATGTTTATAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAATTGACGTTTTTTTTTGGCCCAACATTGCAATTTTTTTTGTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCAAAAGGGCTCCACTACCCCGCGGTCGCCATT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnGlnAlaGlyTyPValThrGlyIleIleGlyLysTrp 138
QY 338 ---TTGGGAAGTCATCCGTCGCCAAAGAAATAGCCATPAACATATCGTTACTGTTTTGG 394
Db 139 HisLeuGlyHisHisGlySerTyPHisProAsnPheArgGly-PheAspTyPTrPheG1 158
QY 395 A-----ACATC 400
Db 158 yIleProTyPSerHisAspMetGlyCysThrAspThrProGlyTyPAsnHisProProCy 178
QY 401 GCCCGTTTCGCC-----GATTCCGCTCAGCGGTATAAAAGAGATCTTTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyPThrAs 198
QY 455 CTGGCTGTCCTTCCATT-----TTAATGTCTTATCTGCTCTCTTGTGATCTTTA 505
Db 198 p---ValAlaLeuProLeuTyPLeuAsnLeuAsnIleValGluGlnProVal----- 214
QY 506 CGGTCTCACTAACCTCTCTTCAACTGCTCAATAAATTTCCCGCTATGCAAAATCCCAAGA 565
Db 215 -----AsnLeuSerSerLeuAlaGln----- 221
QY 566 CTACTTTTACGCTCAAAATCGCTGCGCAACAACAAGCCCTCCACA----- 614
Db 222 ----LysTyPAlaGluLysAlaThrGlnPheIleGlnArgAlaSerGlyArgPr 240
QY 615 ----TTGCGTACCGTGACCATGGCGGAATTTAGAAGGGTGCCTTTGCCACCTATGCTGA 670

Db 240 oPheLeuLeuTyrrValAlaLeuAlaHisMet---HisValProLeu---ProValThrG1 258
QY 671 GGTTCCTATGTTGCTACTCAAACTCCAGCGAGCTCGCTTCTGCTCGGCTTTC 730
Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrrGlyAlaGly----- 272
QY 731 ATTAGAAATGGGAAAAGGAT-----TTGGAGGAGAGACTCAACTCTATCGATCATGA 784
Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACAACAACAATTT 803
Db 290 rValLysGluAsnThrPhe 296

RESULT 3

US-10-121-049-56
; Sequence 56, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-56

Alignment Scores:
Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661A-1 (1-1158) x US-10-121-049-56 (1-525)

QY 32 ACCGCTACTTATTCCTCC-----AAGATCACGTT 61
Db 16 SerGlyPheLeuTyrrProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CTTCTGCCACATATTCCTCCAGCTCCGCGGGTCTTCTGACATTGCT----- 115
Db 36 ProAsnPheValIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGAACTTCCAACTAAGAG----- 136
Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTCTCTTTTTTTTGTAGGCCAATGATAGGAAG-----AACAT 181

Db 75 ValAspPheHisAlaAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAATACGTACAGATATAGTAGATTATGTTTTTATGTTTATAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAATTGACGTTTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGGCTCAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCAAAACGGCTCCACTACCCCGCGCTGCCCATTT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyThrValThrGlyIleIleGlyLysTrp 138
QY 338 ---TTGGGAAGTCATCCGTCCTCCAAAGGAATACCCATACATATCTTACTGTTTGG 394
Db 139 HisLeuGlyHisHisGlySerTyrrHisProAsnPheArgGly-PheAspTyrrTyrrPheG1 158
QY 395 A-----ACATC 400
Db 158 ylleProTyrrSerHisAspMetGlyCysThrAspThrProGlyTyrrAsnHisProProCy 178
QY 401 GCCGCTTTTCGCC-----GATTCGCGCTACGCGGTATATAAAGAGATCTTTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrrThrAs 198
QY 455 CTGGCTGCTCCCTCCATTT-----TTAAATGCTCTTATCTGCTCTCTTGTGATCTTA 505
Db 198 p---ValAlaLeuProLeuTyrrGluAsnLeuAsnIleValGluGlnProVal----- 214
QY 506 CGGCTCACTAACCTCTCTCAACTGCTCAATAATTTCCCGCTATGCAAAATTTCCCAAGA 565
Db 215 -----AsnLeuSerSerLeuAlaGln----- 221
QY 566 CTACTTTTACGCTCAAAATCGCTGCCAACAACAAGCCCTTCCACA----- 614
Db 222 ---LysTyrrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgpr 240
QY 615 ---TTGGCTACCGTGACCATGGCGGAATTTAGAAGGTGCTTTTGCACCTATGCTGA 670
Db 240 oPheLeuLeuTyrrValAlaLeuAlaHisMet---HisValProLeu---ProValThrG1 258
QY 671 GGTTCCTATGTTGCTACTCAAACTCCATGCGAGCTCCGCTTCTGCTCGGCTTCTTC 730
Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrrGlyAlaGly----- 272
QY 731 ATTAGAAATGGGAAAAGGAT-----TTGGAGGAGAGACTCAACTCTATCGATCATGA 784
Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACAACAACAATTT 803
Db 290 rValLysGluAsnThrPhe 296

RESULT 4

US-10-123-904-56
; Sequence 56, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.

Db 36 ProAsnPheValIleIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
Qy 116 -----AAGATACCTCCAACTAAGAG----- 136
Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
Qy 137 -----CATGCTCTCTCTTTTGTGGCCCAATGATAGGAAAG-----AACAA 181
Db 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
Qy 182 AGATTATAATACGTACAGATATAGTATGTTTATGTTAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr----- 104
Qy 242 AATAATTGACGTTTTTTTGGCCCAACATTTGAATTTTTTTTGTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
Qy 302 CCCAAACGGGCTCCACTACCGCGCGCTGCCATT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyThrValThrGlyIleIleGlyLysTrp 138
Qy 338 ---TTGGAGTCTACGCTCCCAAGAAAGTACCATACATATGTTACTCTTTTGG 394
Db 139 HisLeuGlyHisHisGlySerThrHisProAsnPheArgGly-PheAspTyrTyrPheG 158
Qy 395 A-----ACATC 400
Db 158 yileProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
Qy 401 GCCCGTTTCGCC-----GATTCGGCTCAGCGGGTATANAAGAGATCTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
Qy 455 CTGCTGTCTCTCCATT-----TTAATGTCTTATCTCTCTCTCTTTGTGATCTTA 505
Db 198 p-----ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGlnProVal----- 214
Qy 506 CGGTCTCACTCACTCTCTCACTGCTCAATAATTTCCGCTATGCAAAATTTCCCAAGA 565
Db 215 -----AsnLeuSerLeuAlaGln----- 221
Qy 566 CTACTTTAGCTCAAAATCGCTCCCAACACAAAGCCCTTCCACA----- 614
Db 222 ---LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
Qy 615 ---TTGCGTACCTGACCATGCGGGAATTTAGAGGGTGCCTTTGCCACCTATGCTGA 670
Db 240 oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu---ProValThrG 258
Qy 671 GTTCTCTATGTTGTCTACTCAAAACTCCATGGGCGAGCTCGCTTCTGCTCCGCTTCTC 730
Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly----- 272
Qy 731 ATTAGAATGTGGAAAGGAT-----TTGGAGGAGAGACTCAACTCTATCGATCATGA 784
Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
Qy 785 CATGAACAACAACAATTT 803
Db 290 rValLysGluAsnThrPhe 296

RESULT 6

US-10-175-746-56
: Sequence 56, Application US/10175746
: Publication No. US20030027270A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen
: APPLICANT: DeForge, Laura
: APPLICANT: Desnoyers, Luc
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gao, Wei-Qiang

: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Sherwood, Steven
: APPLICANT: Smith, Victoria
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3330R1C353
: CURRENT APPLICATION NUMBER: US/10/175,746
: CURRENT FILING DATE: 2002-06-19
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 550
: SEQ ID NO 56
: LENGTH: 525
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-175-746-56
Alignment Scores:
Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20
US-09-814-661A-1 (1-1158) x US-10-175-746-56 (1-525)
Qy 32 ACCGGCTACTTATTTCCCTC-----AAGATCAGCTT 61
Db 16 SerGlyPheLeuTyrProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
Qy 62 CCTCTCTCCCAACATTCGCGTCCGCGGCGCTCTTCTGACATTTGGT----- 115
Db 36 ProAsnPheValIleIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
Qy 116 -----AAGATACCTTCCCACTAAGAG----- 136
Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
Qy 137 -----CATGCTCTCTCTTTTGTGGCCCAATGATAGGAAAG-----AACAA 181
Db 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
Qy 182 AGATTATAATACGTACAGATATAGTATGTTTATGTTAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr----- 104
Qy 242 AATAATTGACGTTTTTTTGGCCCAACATTTGAATTTTTTTTGTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
Qy 302 CCCAAACGGGCTCCACTACCGCGCGCTGCCATT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyThrValThrGlyIleIleGlyLysTrp 138
Qy 338 ---TTGGAGTCTACGCTCCCAAGAAAGTACCATACATATGTTACTCTTTTGG 394
Db 139 HisLeuGlyHisHisGlySerThrHisProAsnPheArgGly-PheAspTyrTyrPheG 158
Qy 395 A-----ACATC 400
Db 158 yileProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
Qy 401 GCCCGTTTCGCC-----GATTCGGCTCAGCGGGTATANAAGAGATCTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198


```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-921-56

Alignment Scores:
Pred. No.: 0.58      Length: 525
Score: 90.00      Matches: 73
Percent Similarity: 34.25%      Conservative: 39
Best Local Similarity: 22.32%      Mismatches: 100
Query Match: 4.35%      Indels: 115
DB: 9      Gaps: 20

US-09-814-661A-1 (1-1158) x US-10-176-921-56 (1-525)

QY 32 ACCGGCTACTTATCCCC-----AAGGATCACTT 61
DB 16 SerGlyPheLeuTyPrProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CCTTCTGCCAATCATTCCTCGCGAGCTCGCGCGCTCTTCTGACATTTGGT-----115
DB 36 ProAsnPheValIleIleLeuAlaAspAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGAATFACTTCAACTAAGAG-----136
DB 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTCTCTTTTGTAGGCCAATGATAGAAAG-----AACAA 181
DB 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAAATACGTCAGATATAGTATATGTTTATGTTTAGACCTCGTACATAGG 241
DB 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAATTGACGCTTTTGGCCCAACATTTTGAATTTTGTACCTCGCGCTGAG 301
DB 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCCAACCGGCTCCACTACCGCGCGCTGCCCATTT-----337
DB 119 ThrThrLeuAlaGluValGluGlnAlaGlyTyrValThrGlyIleIleGlyLysTrp 138
QY 338 ---TTGGGAAGTCATCGCTCCCAAAAGGAATACCATATCGTTACTCTTTGG 394
DB 139 HisLeuGlyHisHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG 158

; Sequence 56: Application US/10137865
; Publication No. US20030032155A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-865-56

Alignment Scores:
Pred. No.: 0.58      Length: 525
```



```
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661a-1 (1-1158) x US-10-137-865-56 (1-525)

QY 32 ACCGGCTACTTATTCCTCC-----AAGGATCACGTT 61
DB 16 SerGlyPheLeuTyProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CTTCTGCCACATCATTCGCGTCCGACGTCGCGCGCTCTTCTGACATTTGGT-----115
DB 36 ProAsnPheValIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGAATCTTCCAACCTAAGAG-----136
DB 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTTCTCTTTTGTAGGCCAATGATAGAAAG-----AACAA 181
DB 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAATACGTCAGAAATATAGTATATGTTTATGTTTACCTCGCTGAG 241
DB 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 ATATTACGCTTTTGTGCGCAACATTTTGTACCTCGCTGAG 301
DB 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCAAAACGGCTCCACTACCGCGCGTCCGCAATT-----337
DB 119 ThrThrLeuAlaGluValLeuGlnAlaGlyTyrValThrGlyIleGlyLysTrp 138
QY 338 ---TTGGAAGTCATCCGTCGCAAAAGAAATACCCATAATCGTTACTGTTTGG 394
DB 139 HisLeuGlyHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG 158
QY 395 A-----ACATC 400
DB 158 yIleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProC 178
QY 401 GCCGCTTCGCCC-----GATTCGCGCTCAGCGGTATAAAAGAGATCTTTTTC 454
DB 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
QY 455 CTGGCTGCTCCCTCCATTT-----TTAAATGCTTATCTGCTCTTGTGATCTTA 505
DB 198 p---ValAlaLeuProLeuTyProLeuValAsnLeuAsnIleValGluGlnProVal-----214
QY 506 CGGTCTCACTACCTCTCTTCAACTGCTCAATAATTTCCCGGTATGCAAAATTCACA 565
DB 215 -----AsnLeuSerSerLeuAlaGln-----221
QY 566 CTACTTTTACGCTCAAAATCGCTGCAACCAACCAAGCCCTCCACA-----614
DB 222 ----LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
QY 615 ----TTGCGTACCGTGACCATGGCGAAATTTAGAAGGTGCGCTTTGCCACCTATGGCTGA 670
DB 240 oPheLeuLeuTyProValAlaLeuAlaHisMet---HisValProLeu---ProValThrG 258
QY 671 GTTTCTTGTGTTCTACTCAAACTCCATGGGAGCTCCGGTCTTCCGCTCGCTTCCT 730
DB 258 nLeuProAlaLapProArgGlyArgSerLeuTyProGlyAlaGly-----272
QY 731 ATTAGAAATGTGGGAAAGGAT-----TTGAGGAGAGAGACTCAACTCTATCGATCATCA 784
DB 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACACAACAAAATTT 803

Db 290 rValLysGluAsnThrPhe 296
RESULT 10
US-10-140-474-56
; Sequence 56, Application US/10140474
; Publication No. US20030032156A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-474-56

Alignment Scores:
Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661a-1 (1-1158) x US-10-140-474-56 (1-525)

QY 32 ACCGGCTACTTATTCCTCC-----AAGGATCACGTT 61
DB 16 SerGlyPheLeuTyProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CTTCTGCCACATCATTCGCGTCCGACGTCGCGCGCTCTTCTGACATTTGGT-----115
DB 36 ProAsnPheValIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGAATCTTCCAACCTAAGAG-----136
DB 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTTCTCTTTTGTAGGCCAATGATAGAAAG-----AACAA 181
DB 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAATACGTCAGAAATATAGTATATGTTTATGTTTACCTCGCTGAG 241
DB 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 ATATTACGCTTTTGTGCGCAACATTTTGTACCTCGCTGAG 301
DB 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCAAAACGGCTCCACTACCGCGCGTCCGCAATT-----337
DB 119 ThrThrLeuAlaGluValLeuGlnAlaGlyTyrValThrGlyIleGlyLysTrp 138
QY 338 ---TTGGAAGTCATCCGTCGCAAAAGAAATACCCATAATCGTTACTGTTTGG 394
DB 139 HisLeuGlyHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG 158
QY 395 A-----ACATC 400
DB 158 yIleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProC 178
QY 401 GCCGCTTCGCCC-----GATTCGCGCTCAGCGGTATAAAAGAGATCTTTTTC 454
DB 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
QY 455 CTGGCTGCTCCCTCCATTT-----TTAAATGCTTATCTGCTCTTGTGATCTTA 505
DB 198 p---ValAlaLeuProLeuTyProLeuValAsnLeuAsnIleValGluGlnProVal-----214
QY 506 CGGTCTCACTACCTCTCTTCAACTGCTCAATAATTTCCCGGTATGCAAAATTCACA 565
DB 215 -----AsnLeuSerSerLeuAlaGln-----221
QY 566 CTACTTTTACGCTCAAAATCGCTGCAACCAACCAAGCCCTCCACA-----614
DB 222 ----LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
QY 615 ----TTGCGTACCGTGACCATGGCGAAATTTAGAAGGTGCGCTTTGCCACCTATGGCTGA 670
DB 240 oPheLeuLeuTyProValAlaLeuAlaHisMet---HisValProLeu---ProValThrG 258
QY 671 GTTTCTTGTGTTCTACTCAAACTCCATGGGAGCTCCGGTCTTCCGCTCGCTTCCT 730
DB 258 nLeuProAlaLapProArgGlyArgSerLeuTyProGlyAlaGly-----272
QY 731 ATTAGAAATGTGGGAAAGGAT-----TTGAGGAGAGAGACTCAACTCTATCGATCATCA 784
DB 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACACAACAAAATTT 803
```

```
Db 119 ThrThrLeuAlaGluValLeuGlnGlnAlaGlyTyrValThrGlyIleIleGlyLysTrp 138
QY 338 ---TTGGGAGTCATCCGTCACAAAGAAATAGCCATAACATATCGTTACTGTTTGG 394
Db 139 HisLeuGlyHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG1 158
QY 395 A-----AGATC 400
Db 158 ylleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProcy 178
QY 401 GCCCGTTTCGCC-----GATTCCGCCCTCAGCGGTATATAAAGAGATCTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
QY 455 CTGGCTGTCCTTCCATTT-----TTAAATGTCTTATCTGCTCTCTTGTGTCCTTA 505
Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGluGlnProVal----- 214
QY 506 CGGTCTCACTACCTCTCTCAACTGCTCAATAATTTCCCGCTATGCCAAATTTCCCAAGA 565
Db 215 -----AsnLeuSerLeuAlaGln----- 221
QY 566 CTACTTTTACGCTCAAAATCGCTGCCAACACAAAGCCCTTCCACA----- 614
Db 222 ---LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
QY 615 ---TTGGTACCGTACCATGCGGAATTTAGAAAGGTGCTTTGCCACCTATGGCTGA 670
Db 240 oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu---ProValThrG1 258
QY 671 GGTTCCTATGTTGTTACTCAAACTCCATGGCAGCTCCGCTTCTGCTCGGCTTCTTC 730
Db 258 nleuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly----- 272
QY 731 ATTAGAAATGTGGGAAAGGAT-----TTGAGGAGAGACTCAACTCTATCGATCANGA 784
Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACACACAAATTT 803
Db 290 rValLysGluAsnThrPhe 296
```

RESULT 11

US-10-142-431-56
; Sequence 56, Application US/10142431
; Publication No. US20030036179A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56

; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-56

Alignment Scores:

Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661A-1 (1-1158) x US-10-142-431-56 (1-525)

```
QY 32 ACCGGCTACTTATTCCTCC-----AAGATACACGTT 61
Db 16 SerGlyPheLeuTyrProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CCTCTGCCAACATCATTTCCGCTCGAAGCTCGCGCGTCTTCTGACATGGT----- 115
Db 36 ProAsnPheValIleIleLeuAlaAspAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AGAATACTTCAACTAAGAG----- 136
Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTTCTCTTTTGTAGGCCAATGATAGGAAAG-----AACAT 181
Db 75 ValAspPheHisAlaAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAAATACGTCAGAAATATAGTAGATATGTTTTTATGTTAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAATTCACGTTTTTTTGGCCAAACATTTGAATTTTTTTTGTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCCAAACGGCTCCACTACCCCGCGCGTCCCAT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyTyrValThrGlyIleIleGlyLysTrp 138
QY 338 ---TTGGGAGTCATCCGTCACAAAGAAATAGCCATAACATATCGTTACTGTTTGG 394
Db 139 HisLeuGlyHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG1 158
QY 395 A-----ACATC 400
Db 158 ylleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProcy 178
QY 401 GCCCGTTTCGCC-----GATTCCGCCCTCAGCGGTATATAAAGAGATCTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
QY 455 CTGGCTGTCCTTCCATTT-----TTAAATGTCTTATCTGCTCTCTTGTGTCCTTA 505
Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGluGlnProVal----- 214
QY 506 CGGTCTCACTACCTCTCTCAACTGCTCAATAATTTCCCGCTATGCCAAATTTCCCAAGA 565
Db 215 -----AsnLeuSerLeuAlaGln----- 221
QY 566 CTACTTTTACGCTCAAAATCGCTGCCAACACAAAGCCCTTCCACA----- 614
Db 222 ---LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
QY 615 ---TTCCGTCACCGTACCATGCGGAATTTAGAAAGGTGCTTTGCCACCTATGGCTGA 670
Db 240 oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu---ProValThrG1 258
QY 671 GGTTCCTATGTTGTTACTCAAACTCCATGGCAGCTCCGCTTCTGCTCGGCTTCTTC 730
```

```
Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly----- 272
QY 731 ATTAGAAATGTGGAAAGGAT-----TTGAGGAGGAGAGACTCAACTCTATCGATCATGA 784
Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACAACAACAATTT 803
Db 290 rValLysGluAsnThrPhe 296

RESULT 12
US-10-143-114-56
; Sequence 56, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RLC211
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-143-114-56

Alignment Scores:
Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661a-1 (1-1158) x US-10-143-114-56 (1-525)
QY 32 ACCGGCTACTTATCCCC-----AAGGATCACGTT 61
Db 16 SerGlyPheLeuTyrProLeuValAspPheCysIleSerGlyLysThrArgGlyClnLys 35
QY 62 CCTTCTGCCACATCATTCGGCTGCAAGCTCGCGCGCTTCTTGACATTGGT----- 115
Db 36 ProAsnPheValIleLeuAlaAspAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGAATACTTCAACTAAGAG----- 136
Db 55 TrpAlaGluThrLysAspHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 74
QY 137 -----CATGCTCTCTCTTTT-----TAGGCAATGATAGGAAAG-----AACAA 181
Db 75 ValAspPheHisAlaAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAATACGTGACAGATATAGATATGTTTTTATGTTTAGACCTCGTACATAGG 241
||||| :|||
```

```
Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAAATGTGCTTTTGGCCCAACATTTGAAATTTTGTGTTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCCAAACGGGCTCCACTACCCGCGCGGTCCGCATT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyTyrValThrGlyIleIleGlyLysTrp 138
QY 338 ---TTGGGAAGTCACTCGTCCCAAAAGGAAATAGCCATACATATCGTTACTCTTTTGG 394
Db 139 HisLeuGlyHisHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG 158
QY 395 A-----ACATC 400
Db 158 yIleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
QY 401 GCCGTTTTCGCC-----GATTCGGCTCAGCGGGTATAAAAGAGATCTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
QY 455 CTGGCTGCTCCCTTCCATT-----TTAAATGCTTATCTGCTCTCTTGTGATCTTA 505
Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGluGlnProVal----- 214
QY 506 CGGTCTCACTAACTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATTCCCAAGA 565
Db 215 -----AsnLeuSerSerLeuAlaGln----- 221
QY 566 CTACTTTTACGCTCAAAATCGCTGCCAACACAAAGCCCTTCCACA----- 614
Db 222 ---LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
QY 615 ---TTGCGTACCGTGACCATCGCGGAATTTAGAGGGTGCTTTGCCACCTATGCTGA 670
Db 240 oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu---ProValThrGl 258
QY 671 GGTTCCTATGTTGTTACTACTAAAACCTCCATGGCAGCTCCGCTTCTGCTCCGCTTCTTC 730
Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly----- 272
QY 731 ATTAGAAATGTGGAAAGGAT-----TTGAGGAGGAGAGACTCAACTCTATCGATCATGA 784
Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
QY 785 CATGAACAACAACAATTT 803
Db 290 rValLysGluAsnThrPhe 296

RESULT 13
US-10-140-002-56
; Sequence 56, Application US/10140002
; Publication No. US20030037623A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3330RIC59
 ; CURRENT APPLICATION NUMBER: US/10/140,002
 ; CURRENT FILING DATE: 2002-05-06
 ; PRIOR APPLICATION REMOVED - See Palm or File Wrapper
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 56
 ; LENGTH: 525
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-10-140-002-56

Alignment Scores:
 Pred. No.: 0.58 Length: 525
 Score: 90.00 Matches: 73
 Percent Similarity: 34.25% Conservative: 39
 Best Local Similarity: 22.32% Mismatches: 100
 Query Match: 4.35% Indels: 115
 DB: 9 Gaps: 20

US-09-814-661A-1 (1-1158) x US-10-140-002-56 (1-525)

QY	32	ACCGGCTACTTATCCCC	-----AAGGATCACGTT	61
DB	16	SerGlyPheLeuTyProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys		35
QY	62	CCTTCGCAACATCATTCGCGTCGAACGTCGCGCGGCTCTTCTGACATTGGT	-----	115
DB	36	ProAsnPheValIleLeuAlaAspAspMet	---GlyTrpGlyAspLeuGlyAlaAsn	54
QY	116	-----AAGATACCTCCCACTAAG	-----	136
DB	55	TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe		74
QY	137	-----CATGCTCTCTTTT	TTTGTAGGCCAATGATAGGAAG	181
DB	75	ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly		94
QY	182	AGATTATAATACGTCAGAAATATAGTAGATATGTTTTATGTTAGACCTCGTACATAGG		241
DB	95	ArgLeuGlyLeuArgAsnGlyValThr	-----Arg	104
QY	242	AATAATTCACGTTTTTTTGGCCCAACATTTGAATTTTTTTTGTACCTCGCGCTGG		301
DB	105	Asn	-----PheAlaValThrSerValGlyGly	118
QY	302	CCCAACGGGCTCCACTACCGCGCGGCTGCCATT	-----	337
DB	119	ThrThrLeuAlaGluValLeuGlnAlaGlyTyrValThrGlyIleGlyLysTrp		138
QY	338	---TTGGGAAGTCATCCGTCCTCCCAAGAAATAGCCATAATATCGTTACTGTTTGG		394
DB	139	HisLeuGlyHisHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG		158
QY	395	A	-----ACATC	400
DB	158	YileProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProC		178
QY	401	CGCCGTTTCGCC	-----GATTCCGCTCAGCGGGTATAAAAGAGATCTTTTTTTTC	454
DB	178	sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs		198
QY	455	CTGGCTGTCCTCCATTT	-----TTAAATGTCTTATCTGCTCTCTTGTGATCTTA	505
DB	198	p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGlnProVal	-----	214
QY	506	CGGTCTCACTAACCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATCCCAAGA		565
DB	215	-----AsnLeuSerLeuAlaGln	-----	221
QY	566	CTACTTTTACGCTCAAAATCGCTCCCAACCAACAAAGCCCTTCCACA	-----	614
DB	222	---LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr		240

QY	615	---TTGCGTACCGTGACCATGGCGGAATTTAGAGAGGCTGCTTTGCCACCTATCGCTCA	670	
DB	240	oPheLeuLeuTyrValAlaLeuAlaHisMet	---HisValProLeu	258
QY	671	GGTTCCTATGTGCTCTACTCAAACTCCATGGCAGCTCCGCTTCTGCTCCGCTTCCTC	730	
DB	258	nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly	272	
QY	731	ATTAGAAATGTGGGAAAAGGAT	-----TTGAGAGAGAGACTCAACTCTATCGATCATCA	784
DB	273	-----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisThr	290	
QY	785	CATGAACAACAACAATTT	803	
DB	290	rValLysGluAsnThrPhe	296	

RESULT 14

US-10-142-419-56
 ; Sequence 56, Application US/10142419
 ; Publication No. US20030044945A1
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Collin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330RIC244
 ; CURRENT APPLICATION NUMBER: US/10/142,419
 ; CURRENT FILING DATE: 2002-05-10
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 56
 ; LENGTH: 525
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-10-142-419-56

Alignment Scores:
 Pred. No.: 0.58 Length: 525
 Score: 90.00 Matches: 73
 Percent Similarity: 34.25% Conservative: 39
 Best Local Similarity: 22.32% Mismatches: 100
 Query Match: 4.35% Indels: 115
 DB: 9 Gaps: 20

US-09-814-661A-1 (1-1158) x US-10-142-419-56 (1-525)

QY	32	ACCGGCTACTTATCCCC	-----AAGGATCACGTT	61
DB	16	SerGlyPheLeuTyProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys		35
QY	62	CCTTCCTCCCAACATCATTCGCGTCGAACGTCGCGCGGCTCTTCTGACATTGGT	-----	115
DB	36	ProAsnPheValIleLeuAlaAspAspMet	---GlyTrpGlyAspLeuGlyAlaAsn	54
QY	116	-----AAGATACCTCCCACTAAG	-----	136
DB	55	TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe		74

```
QY 137 -----CATGCTCTCTCTTTTGTAGGCAATGATAGGAAAG-----AACAA 181
Db 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAAATACGTACAGATATAGTAGATATGTTTTATGTTAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAATTCACGTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCCAAACGGGTCCACTACCGCGCGGTCCGCCATT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyThrValThrGlyIleLeuGlyLysTrp 138
QY 338 ---TTGGGAAGTCATCCGTCACCAAGGAATACCATATCGTTACTGTTTGG 394
Db 139 HisLeuGlyHisHisGlySerThrHisProAsnPheArgGly-PheAspTyrThrPheG 158
QY 395 A-----ACATC 400
Db 158 ylleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
QY 401 GCCCGTTTCGCC-----GATTCGCGCTCAGCGGTATATAAGAGATCTTTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuArgAspCysTyrThrAs 198
QY 455 CTGGCTGTCCCTCCATT-----TTAAATGCTTATCTGCTCCTTCTGATCCTTA 505
Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGluGlnProVal----- 214

; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C38
; CURRENT APPLICATION NUMBER: US/10/123,262
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 56
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-262-56

Alignment Scores:
Pred. No.: 0.58 Length: 525
Score: 90.00 Matches: 73
Percent Similarity: 34.25% Conservative: 39
Best Local Similarity: 22.32% Mismatches: 100
Query Match: 4.35% Indels: 115
DB: 9 Gaps: 20

US-09-814-661a-1 (1-1158) x US-10-123-262-56 (1-525)
QY 32 ACCGGCTACTATTATCCCC-----AAGATACACGTT 61
Db 16 SerGlyPheLeuTyrProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
QY 62 CCTCTCGCAACATCATTCGCGTCAACGTCGCGCGCTCTTCGACATGGT----- 115
Db 36 ProAsnPheValIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
QY 116 -----AAGATACTTCCAACATAAGAG----- 136
Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
QY 137 -----CATGCTCTCTTTTGTAGGCAATGATAGGAAAG-----AACAA 181
Db 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
QY 182 AGATTATAAATACGTACAGATATAGTAGATATGTTTTATGTTAGACCTCGTACATAGG 241
Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
QY 242 AATAATTCACGTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGCGCTGAG 301
Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
QY 302 CCCAAACGGGTCCACTACCGCGCGGTCCGCCATT----- 337
Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyThrValThrGlyIleLeuGlyLysTrp 138
QY 338 ---TTGGGAAGTCATCCGTCACCAAGGAATACCATATCGTTACTGTTTGG 394
Db 139 HisLeuGlyHisHisGlySerThrHisProAsnPheArgGly-PheAspTyrThrPheG 158
QY 395 A-----ACATC 400
Db 158 ylleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
QY 401 GCCCGTTTCGCC-----GATTCGCGCTCAGCGGTATATAAGAGATCTTTTTTTC 454
Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuArgAspCysTyrThrAs 198
QY 455 CTGGCTGTCCCTCCATT-----TTAAATGCTTATCTGCTCCTTCTGATCCTTA 505
Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGluGlnProVal----- 214
```

RESULT 15

US-10-123-262-56

; Sequence 56, Application US/10123262

; Publication No. US20030049816A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

506	Qy	CGGTCTCAGTAACTCTCTCTTCCAGTGCATCAATAATTTCCCGCTATGCAAAATTCACAAGA	566	Qy	CTACTTTTACGCTCAAAATCGCTGCCACACACACAAGCCCTTCCACA-----	614
215	Db	-----AsnLeuSerLeuAlaGln-----	222	Db	-----LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr	240
566	Qy	CTACTTTTACGCTCAAAATCGCTGCCACACACACAAGCCCTTCCACA-----	615	Qy	-----TTGCGTACCGTGCACCATGGCGGAATTTAGAAGGGTGCTTTGCCACCTATGCCTCA	670
240	Db	oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu-----ProValThrGl	258	Qy	GTGTCTATGTGTCTACTAAAACTCCATGGCAGCTCCGGCTTCTGCCTCCGCTTCTTC	730
566	Qy	CTACTTTTACGCTCAAAATCGCTGCCACACACACAAGCCCTTCCACA-----	614	Qy	CTACTTTTACGCTCAAAATCGCTGCCACACACACAAGCCCTTCCACA-----	670
240	Db	oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu-----ProValThrGl	258	Db	nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly-----	272
731	Qy	ATTAGAAATGTGGAAAAGGAT-----TTGAGGAGAGAGACTCAACTCTATCGATCATCA	784	Qy	ATTAGAAATGTGGAAAAGGAT-----TTGAGGAGAGAGACTCAACTCTATCGATCATCA	784
273	Db	-----LeuTrpGluMetAspSerLeuValGlyClnIleLysAspLysValAspHisTh	290	Db	-----LeuTrpGluMetAspSerLeuValGlyClnIleLysAspLysValAspHisTh	290
785	Qy	CATGAACAACACAAATTT	803	Qy	CATGAACAACACAAATTT	803
290	Db	rValLysGluAsnThrPhe	296	Db	rValLysGluAsnThrPhe	296

Search completed: July 16, 2003, 12:18:33
Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 11:53:05 ; Search time 105 seconds

(without alignments)

4544.811 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 2069

Sequence: 1 aatgagcaaccgtgtcaaca.....cctctcttgatcaatgcttt ll58

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2_1/USPTO.spool/US09814661/runat_16072003_115258_14524/app_query.fasta_1.1351
-DB=SPREMBL_21 -QFMT=fastn -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=dits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09814661.ecgn_1.139 -runat_16072003_115258_14524 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_21:.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	105	5.1	402	8	Q9T6M3

2	104.5	5.1	1411	10	Q9SHP9
3	103.5	5.0	127	4	Q96EF8
4	101.5	4.9	492	8	Q9G8P3
5	101	4.9	307	5	Q96261
6	101	4.9	347	8	Q34096
7	98.5	4.8	221	17	Q8TRD5
8	98	4.7	153	8	Q9T5H1
9	98	4.7	737	10	Q940I5
10	96	4.6	493	5	Q01612
11	96	4.6	302	5	Q97278
12	93.5	4.5	248	8	Q9T238
13	93.5	4.5	355	8	Q34937
14	93	4.5	1411	4	Q96QW6
15	92.5	4.5	508	16	Q98PQ8
16	92.5	4.5	654	3	Q04163
17	92.5	4.5	661	17	Q973L1
18	91	4.4	153	8	Q9T5H3
19	90.5	4.4	531	8	Q47582
20	90	4.3	153	8	Q9T5J4
21	90	4.3	153	8	Q9T5I9
22	90	4.3	153	8	Q9T5I8
23	90	4.3	153	8	Q9T5I7
24	90	4.3	153	8	Q9T5I4
25	90	4.3	153	8	Q9T5I0
26	90	4.3	153	8	Q9T5H6
27	90	4.3	153	8	Q9T5H5
28	90	4.3	153	8	Q9T5H4
29	90	4.3	153	8	Q9T4N7
30	90	4.3	153	8	Q9T4N6
31	90	4.3	153	8	Q9T4N5
32	90	4.3	153	8	Q9T4D5
33	90	4.3	153	8	Q9T3I1
34	90	4.3	501	5	Q93840
35	90	4.3	525	4	Q9Y2K4
36	90	4.3	525	4	Q96EG1
37	89	4.3	153	8	Q99671
38	89	4.3	153	8	Q9T5J7
39	89	4.3	153	8	Q9T5J6
40	89	4.3	153	8	Q9T5J5
41	89	4.3	153	8	Q9T5J1
42	89	4.3	153	8	Q9T5J0
43	89	4.3	153	8	Q9T5I1
44	89	4.3	153	8	Q9T5H9
45	89	4.3	153	8	Q9T5H7

ALIGNMENTS

RESULT 1

ID	Q9T6M3	PRELIMINARY;	PRT;	402 AA.
AC	Q9T6M3;			
DT	01-MAY-2000	(TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DE	NADH-ubiquinone oxidoreductase subunit 4.			
GN	NB4.			
OS	Globodera pallida.			
OC	Mitochondrion.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;			
OC	Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.			
ON	NCBI_TaxID=36090;			
RN	[1]			

SEQUENCE FROM N.A.
MEDLINE=20034761; PubMed=10628979;
Armstrong M.R., Blok V.C., Phillips M.S.;
"A multipartite mitochondrial genome in the Potato Cyst Nematode
Globodera pallida."
Genetics 154:181-192(2000).
-1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
EMBL; AJ249395; CAB55373.1; -
InterPro; IPR001750; Oxidored_g1.
Pfam; PF00361; oxidored_g1; 1.

[illegible]

```

Alignment Scores:
Pred. No.:          0.0383          Length:      127
Score:             103.50          Matches:     14
Percent Similarity: 86.67%        Conservative: 12
Best Local Similarity: 46.67%     Mismatches:  3
Query Match:       5.00%         Indels:      1
DB:                4            Gaps:        1

US-09-814-661A-1 (1-1158) x Q96EF8 (1-127)

QY      994  TGCATTTCCTTCCCTTATATACATATATATATATATATATATATCTCTTC--- 1050
      |||:||||:  |||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      80  CysValtyrIleargValtyrValtyrIleTyrrValtyrIleTyrrValCysMetTyrr 99

QY      1051 TAGGTATTTTGTATTCTCTGTCCTTATC 1080
      |||||  |||:  |||:||||:||||:
Db      100  TyrrValCysIleTyrrCysValtyrIle 109

RESULT 4
Q9G8P3
ID      Q9G8P3          PRELIMINARY;          PRT;      492 AA.
AC      Q9G8P3; 2001 (TEMBLrel. 16, Created)
DT      01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT      01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT      01-MAR-2002 (TEMBLrel. 20, Last annotation update)

```

```

DE NADH dehydrogenase subunit 2 (EC 1.6.5.3).
GN NAD2.
OS Naegleria gruberi.
OC Mitochondrion.
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX NCBI_TaxID=5762;
[1]
SEQUENCE FROM N.A.
RA Burger G., Lang B.F., Nerad T.A., Gray M.W.;
RT "The mitochondrial genome of the supposedly primitive protist,
   Naegleria gruberi.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL: AF288092; AAG17811.1; -.
DR InterPro; IPR001750; Oxidored_q1.
DR DR Pfam; PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 492 AA; 58465 MW; B2A5F3B7119756FF CRC64;

Alignment Scores:
Pred. No.: 0.0754 Length: 492
Score: 101.50 Matches: 80
Percent Similarity: 37.61% Conservative: 49
Best Local Similarity: 23.32% Mismatches: 96
Query Match: 4.91% Indels: 119
DB: 8 Gaps: 19

US-09-814-661A-1 (1-1158) x O9G8P3 (1-492)

```

US-09-814-661A-1 (1-1158) x Q9GBP3 (1-492)

QY	216	TTTTATGTTTAGACCTCTACATAGGAATAATTGACGTTTTTTTTTTGGCCACATTTGAA	275
Db	7	PhelLeuLeuProGluAsnTyrLeu---LeuPheSerIlePhePheIlePheIleTyrLeu	25
QY	276	ATTTTTTTTCTACTCGCGCTGAGCCAAACAGGGCTCCACTACCCGCGGGTCGCCA	335
Db	26	IlePhePheSerPheSerAlaGluLys-----LysPheProLysIysAsnLys	41
QY	336	TTTTGGGAAGTCATCCGTCGCCAAAAGGAATAGCCATAACA-----	377
Db	42	PheThrGluTyrIle-----valLysLeuIleLeuIleAsnPhePheLeuThrLeu	59
QY	378	-----TATCGTTTACTGTTTGGAAACATCGCCGCTT---TCGCCGATATCCGCCTCAGCG	428
Db	60	SerPhePheSerTyrAsnTyrGlyValargAspIlePhePheArgAspSerLeuSerAsn	79
QY	429	GGTATAAAGAGAT-----CTTTTTTTTCTGGCTGCTCCCTTCATTTTAA	476
Db	80	GlyIleGlnLeuIleAsnIleIleLeuValPheLeuPhePheThrCysValProPheLeu	99
QY	477	-----AATGCTTTATCTGCTCCTTTGTCATCT	503
Db	100	LeuAsnLysTyrSerLeuSerSerPheGluPheThrIleLeu-IleLeuLeuCysLeuIle	119
QY	504	TACGGTCTCACTAACTCTCTCAACTGCTCAATAATTTCCGCGTATCCAAAATCCCAA	563
Db	119	uSerLeuAsnLeuLeu-ValIleThrValAsnLeuIleSerPheTyr-----	134
QY	564	GACTACTTTTACGCTCAAAATCGTGCCACACAAACAGCCCTTCCACATTGCGTACC	623
Db	135	--LeuLeuLeuGluPheGlnSerIle-----CysPheTyrIle-----	146
QY	624	GTGACCATGCGGAATTTAGAAGGGTGCTTTGGCCACCTATCGCTCAGGTTCTCATGTG	683
Db	147	-----LeuAlaSerTyrGly-----	151
QY	684	TCTACTCAAAACTCCATGGGAGCTCCGCTTCTGCTCGCTCTTCTTATTAGAAATGTGG	743
Db	152	----LysLysAsnLysTyrSerPheGluAlaGlyLeuLysTyrPheIleLeuGlySerP	170
QY	744	GAAGAAG-----TTTGGAGGAGAGACTCAACTCTATTC	776
Db	170	heSerSerValLeuLeuLeuPheGlyIleAlaLeuPheTyrGlyPheThrGlyIlePheP	190

OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Smirnov S., Atnoor D., Brown A.,
RA FitzHugh W., Calvo S., Engels R., Dearellano K., Johnson R.,
RA Linton L., McEwan P., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., Stange-Thomann N., Dearellano K., Johnson R.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hederich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010710; AAM03946.1;
KW Complete proteome.
SQ SEQUENCE 221 AA; 25774 MW; 7FCE996491D4859B CRC64;

Alignment Scores:
Pred. No.: 0.142 Length: 221
Score: 98.50 Matches: 41
Percent Similarity: 38.85% Conservative: 20
Best Local Similarity: 26.11% Mismatches: 64
Query Match: 4.76% Indels: 32
DB: 17 Gaps: 5

US-09-814-661a-1 (1-1158) x Q8TDS (1-221)
Qy 661 CTATGGTGGAGTTCCTATGTTGT-----CTACTAAACATCCATGGCAGCTCCG 711
Db 23 MetTrpIleGluPheIleCysAsnPheIleThrLeuHisMetProTyrArgGluPhe 42
Qy 712 CTTCTGCTCCGCTTCCTCATAGAAATGCGGAAGAGGATTCGGAGGAGAGACTCACT 771
Db 43 IleAsnProLeuMetLeuTyr-----Leu 50
Qy 772 CTATCGATGATGACATGAACAAACAAATTTGGTTCTGGCGAACTAAATCTATGTCA 831
Db 51 LeuSerLeuPheLeuProAlaAlaGlyLeuValPheMetLysAsnAsn----- 67
Qy 832 ACCAGGTAAGTGCAGGAATGAGCTTCTAAAGTCTCCTTTCATCTC----- 879
Db 68 --ArgLysIleSerSerArgTrpTyrLeuGlyIleProPheLeuLeuGlyLeuLeu 86
Qy 880 -----TTTTCTTCTCTCTTCCATTCCTCCACTAGTTCGTGTTCTTTC 921
Db 87 AsnThrPheLeuArgAspPheMetPheMetAsnProPheSerProTyrAlaValAsnLeu 106
Qy 922 TTCCTTAGATACCCCTCTTTTTCAGGAGCTCTCGCTCTACTATGTTGCTCATTCGAAA 981
Db 107 PheTyrGlyGlnProPheIleAlaThrPheProIleLeu-TyrArgTrpHisIleLeuG 126
Qy 982 CATCTCTCCGTCGATTTTCCTTT-----CCCTTTATACATATATATATATATA 1035
Db 126 uPheLeuSerProAsnAspSerIleValAsnSerThrValLeuPhePheIleTyrIlePh 146
Qy 1036 TATATATGCTCTCTCTAGCTATTTTGTATTTCTGCTCTCTTATCAAG 1084
Db 146 ePheSerIleSerValAlaLeuPheCysLeuSerValProGluGluLys 162

RESULT 8
Q9T5H1
ID Q9T5H1 PRELIMINARY; PRT; 153 AA.
AC Q9T5H1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
OC NADH dehydrogenase subunit 4 (Fragment).
GN NDA.
OS Mazamastrongylus odocoilei.
OG Mitochondrion.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Ostertagiinae; Mazamastrongylus.
OX NCBI_TaxID=85880;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOTS05;
RX MEDLINE=99083438; PubMed=9866206;
RA Blouin M.S., Yowell C.A., Courtney C.H., Dame J.B.;
RT "Substitution bias, rapid saturation, and the use of mtDNA for
RT nematode systematics.";
RL Mol. Biol. Evol. 15:1719-1727(1998).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; AF070872; AAC99741.1;
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; Oxidored_q1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
FT NON_TER 1
SQ SEQUENCE 153 AA; 17804 MW; 21E6A7FD7A314495 CRC64;

Alignment Scores:
Pred. No.: 0.153 Length: 153
Score: 98.00 Matches: 28
Percent Similarity: 45.83% Conservative: 16
Best Local Similarity: 29.17% Mismatches: 34
Query Match: 4.74% Indels: 18
DB: 8 Gaps: 2

US-09-814-661a-1 (1-1158) x Q9T5H1 (1-153)
Qy 866 TTCCTTTCATACACTCTTTCTTTCTTCTTCCATTTCCACATGCTCTGTTCTTTCTTCT 925
Db 53 PheMetAsnSerSerMetIlePheSerIlePhePheProLeuIlePheLeuSerAsnSer 72
Qy 926 CTTAGATACCCCTCTTTTCAGGAGCTCTCGCTCTACTATTGTTGTCATCTCGAACAATT 985
Db 73 GlyValProProSerLeuSerPheLeuSerGluPheMetIleLeuLeuAsnSerMetIle 92
Qy 986 CTCCTCCGTCGATTTTCTTTCCCTTTATATACATATATATATATATATATATATATAT 1045
Db 93 MetSerLysLeuPhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 111
Qy 1046 TCT-----TCTACG 1054
Db 112 SerLeuPheLeuIleValCysSerPheAlaGlyLysSerPheIleAspTyrAsnAsnTrp 131
Qy 1055 TATTTTGTATTTCTGCTCTTTTATCAAGATAGTCTATATACGTTT 1102
Db 132 AsnPheSerValSerValSerLeuIleMetMetPheAsnIlePhe 147

RESULT 9
Q940I5
ID Q940I5 PRELIMINARY; PRT; 737 AA.
AC Q940I5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 82.5 kDa protein.
GN 297341.3
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

```

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.,
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AV054590; AAK96781.1;
DR InterPro: IPR004813; ISP4_OPT;
DR InterPro: IPR004848; OPT_supfam;
DR Pfam: PF03169; OPT; 1;
DR TIGRFAMs: TIGR00727; ISP4_OPT; 1;
DR TIGRFAMs: TIGR00728; OPT_sfam; 1;
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1;
KW Hypothetical protein.
SQ SEQUENCE 737 AA; 82470 MW; 306438E41E7755EA CRC64;

Alignment Scores:
Pred. No.: 0.189 Length: 737
Score: 98.00 Matches: 62
Percent Similarity: 36.04% Conservative: 49
Best Local Similarity: 20.13% Mismatches: 83
Query Match: 4.74% Indels: 114
DB: 10 Gaps: 16

US-09-814-661a-1 (1-1158) x Q94015 (1-737)
QY .440 AGATCTTTTTCCTGGCTGCTCCCTTCCATTTTAAATGCTTATGCTGCTCTTG 499
DB |||:|||||
DB 44 ArgThrTrpPheLeuGlyLeuThrSerCysValLeuLeuLeuPheLeuAsnThrPhe 63
QY 500 ATCTACGGCTCCTACCTACCTCTCTTCACTGCTCATAATATTTCCCGCTATGCAAAATTC 559
DB |||:|||||
DB 64 ThrTyrArgThrGln---ProLeuThrIleSerAlaIleLeuMetGlnIleAlaValLeu 82
QY 560 CCAAGACTACTTTTACGCTCAAAATCGCTGCTCAACCAACAACA----- 601
DB |||:|||||
DB 83 ProIleGlyLysPheMetAlaArgThrLeuProThrThrSerHisAsnLeuLeuGlyTrp 102
QY 602 -----AGCCCTTCCACATTCGCTACCGTGACCAT----- 631
DB |||:|||||
DB 103 SerPheSerLeuAsnProGlyProPheAsnIle-----LysGluHisValIleIleThr 120
QY 632 -----GCCGGAATTTAGAGGCTGCTTTCGCCACTATGCTGAGGTTCTCAT 679
DB |||:|||||
DB 121 IlePheAlaAsnCysGlyValAlaTyrGlyGlyGlyAspAlaTyrSerIleGlyAlaIle 140
QY 680 GTTCTCTACTCAAACTCCATGGGCAG---CTCGGCTTCTGCTCCCTCTTCTCAT 733
DB |||:|||||
DB 141 ThrValMetLysAlaTyrTyrLysGlnSerLeuSerPheIleCysGlyLeuPheIleVal 160
QY 734 -----AGAAATGTGGGAAGGATTTGGAGGA----- 760
DB |||:|||||
DB 161 LeuThrThrGlnIleLeuGlyTyrGlyTrpAlaGlyIleLeuArgArgTyrLeuValAsp 180
QY 761 -----CAGACTCACTCTATCCATCATGACAT 787
DB |||:|||||
DB 181 ProValAspMetTrpTrpProSerAsnLeuAlaGlnValSerLeuPheArgAlaLeuHis 200
QY 788 GAACAACAACAATAATTTGGTCTGCGCACTAAATCTATCTTCAACACGAGGTGAAGTCGA 847
DB |||:|||||
DB 201 GluLysGlu-----AsnLys 205
QY 848 GGAATGGACTT-----CTAAGATTC-----CTTTCATCTCTTTCTTTTC 889
DB |||:|||||
DB 206 SerLysGlyLeuThrArgMetLysPhePheLeuValAlaLeuGlyAlaSerPheIleTyr 225
QY 890 -----TCTTTCCATTTCCCACTAGTCTGTTCTTTCTCTCTCT 931
DB |||:|||||
DB 226 TyrAlaLeuProGlyTyrLeuPheProIleLeuThrPhePheSerTrpValCysTrpAla 245
QY 932 TACCCCTTCT----- 940
DB |||:|||||
DB 246 TrpProAsnSerIleThrAlaGlnGlnValGlySerGlyTyrHisGlyLeuGlyValGly 265

```

```

QY 941 -----TTTCAGGACTCTGCTCTACTAT-----TGTTCAT 973
DB |||:|||||
DB 266 AlaPheThrLeuAspTrpAlaGlyIleSerAlaTyrHisGlySerProLeuValAlaPro 285
QY 974 TCTCGAAACATTTCTCCCGTGCATTTTCTTTCCCTTTATATATATATATATATATA 1033
DB |||:|||||
DB 286 TrpSerSerIleLeuAsnValGlyValGlyPhe-----IleMetPheIle 300
QY 1034 TATATATATGCTCTCTTCTACGATTTTGTATTTTCTGTGCTGCTTTATCAAGATAGTCTCAT 1093
DB |||:|||||
DB 301 TyrIleIleValProValCysTyrTrp-----LysPhe 311
QY 1094 AATACGTTTGATACAGCTAGATAT 1117
DB |||:|||||
DB 312 AsnThrPheAspAlaArgLysPhe 319

```

RESULT 10

```

O01612 ID O01612 PRELIMINARY; PRT; 493 AA.
AC O01612;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 57.0 kDa protein.
GN T19H12.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Davidson S.;
RT "The sequence of C. elegans cosmid T19H12.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: U97009; AAC69036.1; -.
DR HSP; P20393; IAGY.
DR InterPro: IPR000536; Hormone_rec_lig.
DR Pfam: PF001628; Znf_C4steroid.
DR PRINTS: PF00105; zf-C4; 2.
DR PRODOM: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein; Receptor;
KW Transcription regulation; Zinc-finger.
SQ SEQUENCE 493 AA; 57004 MW; 743578EB1CF8460A CRC64;

```

Alignment Scores:

```

Pred. No.: 0.294 Length: 493
Score: 96.00 Matches: 43
Percent Similarity: 43.66% Conservative: 19
Best Local Similarity: 30.28% Mismatches: 58
Query Match: 4.04% Indels: 22
DB: 5 Gaps: 6

```

US-09-814-661A-1 (1-1158) x 001612 (1-493)

QY 360 AAGAAATAGCCATACATATGTTTGGAAATCGCCGCTTTCGCCGCGATTCC 419
 Db 20 LysArgIleAspCysAsnLeuArgValAsnPheGlyThr-Thr-ProphePheGluTyTh 39
 QY 420 GCTCAGCGGGTATAAAGAGATC---TTTTTTTCTGCTGCTGCTCCATTTTAA 476
 Db 39 rProCysGlnIleProValGlnIleSerPhePheLeuLeuCysProPhe----- 56
 QY 477 AATGCTTATCGCTCTGCTGATCTTACGCTCTCACTAACTCTCTCACTGCTCAA 536
 Db 57 -AlaSerTyHisLeuPheCys-CysLeuArg----- 66
 QY 537 TAATTTCCCGCTATGCAAAATTC---CCAGACTACTTTTACGCTCAAAATCGCTGCCAA 593
 Db 67 --LeuSerAspTyHisThrLysProAlaLeuLeuPheAsnSerLysThrMetThL 86
 QY 594 CAACAACAAGCCCTTCCACATGCTGACGCTGACCATGGCGAATTTAGAGGGTGCCT 653
 Db 86 ysileLeuPheProCysArgIleCysGlyLysLysAlaHisGlyThrHisPheGlyValP 106
 QY 654 TTGCCACCTATGCTGAGGTCTCTAT-----GTTGCTCTACTCAAAACTCCATGGG--- 703
 Db 106 heSerCysArgAlaCysAlaAlaPheArgLeuValPheSerLeuIleLeuGlyMeta 126
 QY 704 -----CAGTCGCGCTCTCCCTCCGCTTCTCATTAGAAATGTGGAAAGGATTTGGAG 758
 Db 126 laileSerPheArgSerCysMetGluIlePheValLysCysLeuAlaLysGlyGlySerG 146
 QY 759 GA 760
 Db 146 ly 146

RESULT 11

QY 097278 PRELIMINARY; PRT; 3032 AA.
 AC 097278;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 363.6 kDa protein.
 GN PF0875W, MAL377.4.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
 RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moulé S.,
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
 falciparum";
 RL Nature 400:532-538 (1999).
 DR EMBL: AL034559; CAB39024.2; -
 DR InterPro: IPR003439; ABC_Transportr.
 DR Pfam: PF00005; ABC_tran; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 3032 AA; 363590 MW; C993891AC27BCE35 CRC64;

Alignment Scores:
 Pred. No.: 0.375 Length: 3032
 Score: 96.00 Matches: 49
 Percent Similarity: 41.81% Conservative: 25
 Best Local Similarity: 27.68% Mismatches: 55
 Query Match: 4.64% Indels: 49
 DB: 5 Gaps: 11

US-09-814-661A-1 (1-1158) x 097278 (1-3032)

QY 581 AAATCGCTCCCAACAACAAGCCCTTCCACATTGCTGCTGACCATGCGCGAATT 640
 Db 1888 LysArgIleAspThrLeuAsnGluProPheAsnIleLysPheHisGluHis----- 1904
 QY 641 TAGAAGGGTGCCTTTGCCACCTATGCTGAGGTTCCTATGTT---GTCCTACTCAAAATC 697
 Db 1905 -----PheMetArgAspPheTyHisLeuAsnIleTyHisValPheLeu 1917
 QY 698 CATGGCAGCTCGGCTTTCGCTTCTTCATTAGAAATGTGGGAAAGATTGGA 757
 Db 1918 SerIleValIlePhePheCysVal---PhePheGluArg-----PheLys 1931
 QY 758 GGAGAGACTCAA-----CTCATGATCATGACATGAACAACAACAATTTGGTTC 808
 Db 1932 AsnGluIleGluAsnArgLysIlePheGluAsnPheHisValHisGlnTyHisLys 1951
 QY 809 TGGCGAACTAAATCTATGTTCAACACAGGTAAGGTGAGGAAATGGACTTCTAAAGTTC 868
 Db 1952 PheGlnIleLeu-----LeuLeuGluTy 1959
 QY 869 CTTTCATATC-----TTTTCTTTCTCTTTCCATTCCCATAGT 909
 Db 1960 LeuTyTyTyPheIleTyHisLeuCysLeuPheIleValLeuTyHisPheAspTyLys 1979
 QY 910 TCTGTTCTTTCTCTCTTAGATACCTTCTTTTCAGGACTCTCGCTCTACTATTGTTG 969
 Db 1980 GluPheLeuPheMetSer-----PhePhe-----CysPheLeuLeuTy 1993
 QY 970 TCATTCTCGAAACATTTCTCTCCGCTGCTTTCTTTCCCTTTATATATATATATA 1029
 Db 1994 GlyPheAsnIlePheLeuSerIleCysLeuPheSerSerLeuTyHisLysTyHis 2013
 QY 1030 TATATATATATGCTCTCTCTTCTACGATTTTCTGTTCTGCTCTTTATC 1080
 Db 2014 -----LeuPheLeuPhePheAsnPheIle---PheCysGlyIle 2026

RESULT 12

QY 097238 PRELIMINARY; PRT; 248 AA.
 ID 097238
 AC 097238;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE SECY-independent transporter protein.
 GN YNF16.
 OS Phytophthora infestans (Potato late blight fungus).
 OG Mitochondrion.
 OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 OC Phytophthora.
 OX NCBI_TaxID=4787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WEST VIRGINIA 4;
 RA Lang B.F., Forget L.;
 RT "The mitochondrial genome of Phytophthora infestans";
 RL (In) O'Brien S.J. (eds.);
 RL Genetic Maps, pp.3:133-135, Cold Spring Harbor Laboratory Press,
 NY (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WEST VIRGINIA 4;
 RA Paquin B., Roewer I., Wang Z., Lang B.F.;
 RT "A robust fungal phylogeny using the mitochondrially encoded nad5
 protein sequence";
 RL Can. J. Bot. 73:S180-S185(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WEST VIRGINIA 4;
 RX MEDLINE=97306266; PubMed=9162109;
 RA Paquin B., Laforest M.J., Forget L., Roewer I., Wang Z., Longcore J.,

Best Local Similarity: 25.43%		Matches: 108	
Query Match: 4.49%		Indels: 81	
DB: 4		Gaps: 13	
US-09-814-661A-1 (1-1158) x Q960W6 (1-1411)			
QY	293	CCGCGTGAGCCCAACAGGGCTCCACTACCGCGCGGTGCGCATTTTGGGAAGTCATACCG	352
DB	692	LysLeuArgProLysArgAlaSerAlaGluGlnSerValLeuPheLysSerLeuHisThr	711
QY	353	TCCCAAAAGGAATAGCATAACATATCGTTACTGTTTGGAAACATCCCGCTTTTCGCC	412
DB	712	AsnThrAsnGlySerLeuProLeuValMetProGluIleAsn-AspLysGluAsnAr	731
QY	413	CGATTCCGCTCAGCGGGTATAAAAGAGAT	454
DB	731	gAspValThrAsnGlyGlyIleLysArgSerArgLeuGluLysSerAlaLeuPheSerSe	751
QY	455	CTGGCTG---TCCCTTCCATTTTAAATGCTTATCTGCTCTTTGTGATCTTACGGTCT	511
DB	751	rLeuLeuSerSerLeuProGlnAspLysIlePheSerProSerValThrSer-	768
QY	512	CACCTACCTCTCTCAACTGCTCAATAATTTCCCGCTATGCCAAATTCCTCAAGACTACT	571
DB	769	-ValAsnThrMetThr-----AlaPheSerThrSerGlnAsnGlySe	783
QY	572	TTACGCTCAAAATCGCTGCCAACAACAGCCCTTCCACATTCGCTACCGTGACCAT	631
DB	783	rLeuSerGlnSerValSerGln-----ProThrThr-----	794
QY	632	GCGGAATTTAGAGGGTGCCTTTGCCACCTATGGCT-----GAGGTTCCTATGTT	682
DB	795	-----GluGlyAlaProProCysGlyLeuAsnLysGluGlnSerAsnLe	809
QY	683	GTCTACTCAAACTCCATGGGC-----AGCTCCGCTCTGCTCCGCTCTCTC	730
DB	809	uLeuProAspAsnSerLeuLysValPheAsnPheAsnSerSerSerThrSerHisSerSe	829
QY	731	ATTA-----GAAATCTGGGAAAGGATTT	754
DB	829	rLeuLysSerProSerHisMetGluLysTyrProGlnLysGluLysThrLysGluAspLe	849
QY	755	GGAGGAGACTCAAC-----TCAT	775
DB	849	uAspSerArgSerAsnLeuHisLeuProGluThrLysPheSerGluLeuSerLysLeuLy	869
QY	776	CGATCATGACATGACAACAACAAATTTGGTCTGCGGAACATAAATCTATGTTCAACCA	835
DB	869	sAsnAspAspMetGluLysAlaAsnHisIleGluSerValIleLysSerAsnLeuProAs	889
QY	836	GGTAAGGTGCGAGAAATGGACTTCTAAAGTTCTCTCATACTCTTTTCTCTCTTC	895
DB	889	ncysAlaAsnSerAspThrAspPheMetGlyLeuPheLysSerSerArgTyrAspProSe	909
QY	896	CATTTCCTCCACTAGTCTGTTCTTCTCTTAGATACCTCTCTTTCAGGACTCT--	953
DB	909	rIleSer-----PheSerGlyMetSerLeuSerAspThrMetThrLeuArgGlySerVa	927
QY	954	-----CGTCTACTATTGTTGTCATT-----	974
DB	927	IcInAsnLysLeuAsnProArgProGlyLysValValIleTyrSerGluProAspValSe	947
QY	975	-----CTCGAAACATTTCTCCCGGTG	995
DB	947	rGluLysCysIleGluValPheSerAspIle	957
RESULT 15			
Q98PQ8			
ID	Q98PQ8	PRELIMINARY:	PRT: 508 AA.
AC	Q98PQ8:		
DT	01-OCT-2001 (TREMBLrel. 18, Created)		
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		

RESULT 15

Q98PQ8	
ID	Q98PQ8
AC	Q98PQ8; PRELIMINARY; PRT; 508 AA.
DT	01-OCT-2001 (TREMBLrel. 18, Created)
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE	Amino acid permease.			
GN	MYPU_6610.			
OS	Mycoplasma pulmonis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;			
OC	Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=2107;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=UAB CTIP;			
RX	MEDLINE=21267165; PubMed=11353084;			
RA	Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,			
RA	Moszer I., Dydvig K., Wroblewski H., Viari A., Rocha E.P.C.,			
RA	Blanchard A.;			
RT	"The complete genome sequence of the murine respiratory pathogen			
RT	Mycoplasma pulmonis.";			
RL	Nucleic Acids Res. 29:2145-2153(2001).			
DR	EMBL; AL445565; CAC13834.1; -			
DR	MyPUList; MYP0_6610; -			
DR	InterPro; IPR002293; AA/rel_primeasel.			
KW	Complete proteome.			
SQ	SEQUENCE 508 AA; 57402 MW; 9BF07F690D720481 CRC64;			
Alignment Scores:				
Pred. No.:	0.703			
Score:	92.50			
Percent Similarity:	33.77%			
Best Local Similarity:	19.39%			
Query Match:	4.47%			
DB:	16			
US-09-814-661A-1 (1-1158) x Q98PQ8 (1-508)				
QY	60	TTCTCTTCTGCCAACATCATTC-	-----CCGTCGAACGTCGCGCG	98
DB	97	PhePheIleProValSerLeuPheThrMetThrIleTyrValThrSerSerPheValasp	116	
QY	99	GTCCTTCTGACATTC-	-----GTAAAGTAATCTTCCAACTAAGAGCATGCTTCTC	146
DB	117	ValGlyValSerLeuLysAsnGlyTrpLeuValLeuPheAlaPheLeuIlePheLeu	136	
QY	147	TTTTTTTTTGTAGGCCAATGATAGGAAGACAATAGATTATAATACGTCAAGATATAG	206	
DB	137	PhePheMetVal-	-----IleAsnLeuIleSerIleLysIleGly	149
QY	207	TAGATATGTTTATGTTTATGACCTCGTACATAGGAATAATGACGTTTTCGGCCA	266	
DB	150	GluValAlaGlnTrpIle--ThrThrIleValLysValValProLeuPheValLeuPro	168	
QY	267	ACATTTGAAATTTTTTTTGTACCTCGCTGAGCCCAACGGGCTCCACTACCGCGCG	326	
DB	169	ValIleAlaPheValPheAlaAspLeuGluLeuGlyAsnThrPheLeuGln-	-----185	
QY	327	CGGTCCGCATTTTGGGAAGTCATCCGTCCTCCAAAGGAATAACCATATCGTTAC	386	
DB	186	-----LysGlnIleLysPro-----GluValGlyIleThr	195	
QY	387	TGTTTTTGGAAATCGCCCGCTTTCGCGCGATTCGCGCTCAGCGGTATATAAAGAGATCTT	446	
DB	196	-----GlyMetSer-	-----198	
QY	447	TTTTTTTTTCCGTGCTGCCCTTCCATTTTTAAATGCTTATCTGCTCTCTTGTGATCTTAC	506	
DB	199	-----LysTrpLeu-----IleIleIleAlaGlyLeu-	207	
QY	507	GGTCTCACTAACCTCTCTTCAACTGCTCAATAATTTCCGCTATGCAAAATTCCTCCAGAC	566	
DB	208	-----ProAlaIleThrPheAlaTyrAsp	215	
QY	567	TACTTTTACGCTCAAAATCGTCCCAACAACAGCCCTTCCACATTTGCTACCGTG	626	
DB	216	AsnPheTyrAlaIleSerAsnIleLysGluLeuSerProLysAlaGluLysIle	235	
QY	627	ACCATGGCGGAATTTAGAAAGGGTGCCTTTGCCACCTATGGCTGAGGTTCTCTATGTTGCT	686	

```
Db      236 GlyMetAlaIleVal-----IleGlyLeuAlaIleThrValIleTyrMetLeu 253
QY      687 ACTAAACTCCAGCGAGCTCCGCTTCTGCTCGCTTCTTCAATGAAATGTGGAA 746
Db      254 ThrValAlaPheAsnTyrGlySerAsnAspGlyThrIleAsnThrIleSerPheLeuGlu 273
QY      747 AAGGATTTGGAGGAGACTCAACTCTATCGATCATGAC----- 785
Db      274 LysProGluAsnAlaGluPheAlaPheAsnSerCysIleAlaIleGlyIleLeu 293
QY      786 -----ATGACAAACAACAATTTGGTTGCG-----GACTAAATCTATGTCAACCAG 836
Db      294 GlyIleIleAsnSerLeuAlaLeuSerSerProTyrGlnLeuArgGlyLeuTyrGlnGln 313
QY      837 GCTAAGTCGAGGAATGCACTTCTAAAGTTCCTTT----- 872
Db      314 GlyGluAlaAsnGluPheArgPheLeuHisTyrPheIleTyrLysIleIleLeuLysGln 333
QY      872 ----- 872
Db      334 GluValAspValLysAsnArgLysGlnThrLeuPheValSerTrpIleTyrLeuPheLeu 353
QY      873 ---CATACTCTTTCTTT----- 887
Db      354 SerSerThrLeuPhePheIleValPheGlyLeuIleAlaIleLeuAlaTyrArgIleAsp 373
QY      888 -----TCTCTTCCATTCC 902
Db      374 ThrTrpAspLeuGlyGlnTyrGlyThrGlyThrTyrLeuTyrSerPheValAspValLeu 393
QY      903 CACTAGTCTGTTCTTTTCTTCTC-----TTAGATACCCCTT 938
Db      394 ThrAsnTyrSerSerPheLeuLeuPheSerIleIleThrValThrIleLeuGlyAlaLeu 413
QY      939 CTTTCAGGACTCTCGT-----CCTACTATT 965
Db      414 IleAsnArgLysThrLysLysIleLysThrAlaLysLysTyrPheIleProThrAla 433
QY      966 GTTGTCATTCTCGAAACATTCTCCGCGCATTTTCTTCCCTTTATATACATATATA 1025
Db      434 IleValThrIleValPhePhe-----PheSerThrPheAlaTyrIle 447
QY      1026 TATATATATATATATCTCTCTAGGTATTTTGTATTCTGTCTTTATCAAGA 1085
Db      448 PheVal-----ValSerValAlaAsnIlePheIleLeuGluGlyLysAspGlnGln 464
QY      1086 TAGTCTAATACGTTG---ATACAGCTAGATATCGCTAGCGCCACATGTGCCCC 1139
Db      465 SerSerIleIleLysLeuValIleLeuLeuIleIleLeuValSerIleIlePro 483
```

Search completed: July 16, 2003, 11:57:55
Job time : 119 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 10:57:32 ; Search time 1643 Seconds
(without alignments)
11414.713 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 1158

Sequence: 1 aatgagcaaccgtgtcaaca.....cctctcttgatgaatgcttt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	663.8	57.3	784	17	AQ875860
C 2	549.2	47.4	796	17	AQ875934
C 3	448.4	38.7	535	17	AQ502431
C 4	437	37.7	589	17	AQ873494
C 5	413.8	35.7	590	17	AQ873374
C 6	406.2	35.1	526	17	AQ873357

7	367.4	31.7	573	17	AQ872886	AQ872886 V54A9 mTn
8	366.8	31.7	499	17	AQ873058	AQ873058 V58E5 mTn
9	366.6	31.7	466	17	AQ501881	AQ501881 V13D1 mTn
10	347.4	30.0	562	17	AQ500866	AQ500866 V31A2 mTn
11	234.2	20.2	543	17	AQ500106	AQ500106 V30F8 mTn
12	189	16.3	547	17	AQ873740	AQ873740 V74E5 mTn
13	137.2	11.8	525	17	AQ502957	AQ502957 V47B7 mTn
14	76.4	6.6	543	17	AQ501433	AQ501433 V22C5 mTn
15	74.8	6.5	530	17	AQ503366	AQ503366 V56E10 mT
16	62	5.4	924	17	CNS01GHN	AL143148 Anopheles
17	61.6	5.3	890	17	CNS00AFU	AL055222 Drosophil
18	60.8	5.3	730	17	CNS03FVU	AL242211 Tetraodon
19	60.2	5.2	899	17	CNS00C89	AL058348 Drosophil
20	60.2	5.2	1101	17	CNS012JN	AL101645 Drosophil
21	60	5.2	565	17	CNS03LE2	AL249347 Tetraodon
22	60	5.2	997	17	CNS005TE	AL060767 Drosophil
23	59.8	5.2	792	17	AG041274	AG041274 Pan trogl
24	59	5.1	1256	12	BE966592	BE966592 60160276
25	58.8	5.1	408	17	CNS016NE	AL105964 Drosophil
26	58.8	5.1	1101	17	CNS00LT2	AL078714 Drosophil
27	58.6	5.1	393	17	AQ557440	AQ557440 HS_2081-A
28	58.4	5.0	1101	17	CNS0039R	AL063932 Drosophil
29	57.8	5.0	408	9	AL514793	AL514793 AL514793
30	57.8	5.0	623	17	AQ157491	AQ157491 nbxb0009H
31	57.8	5.0	861	17	AQ026945	AQ026945 CIT-HSP-2
32	57.8	5.0	926	12	BF274178	BF274178 GA_EB002
33	57.6	5.0	430	17	CNS01LNB	AL100481 Drosophil
34	57.6	5.0	448	9	AL513961	AL513961 AL513961
35	57.6	5.0	845	17	CNS0332K	AL225605 Tetraodon
36	57.6	5.0	960	17	AG031602	AG031602 Pan trogl
37	57.4	5.0	1101	17	CNS0006J	AL062049 Drosophil
38	57.2	4.9	886	17	CNS0075DF	AL429961 clone BA0
39	57.2	4.9	991	13	BM414939	BM414939 OP20007 M
40	57	4.9	476	17	CNS04TQC	AL306813 Tetraodon
41	57	4.9	761	17	AG152484	AG152484 Pan trogl
42	57	4.9	887	17	CNS060V4	AL408566 T3 end of
43	57	4.9	1101	17	CNS0160P	AL107011 Drosophil
44	56.8	4.9	620	17	AQ158266	AQ158266 nbxb0010N
45	56.8	4.9	927	13	BM415589	BM415589 OP20666 M

ALIGNMENTS

RESULT 1	AQ875860/c	AQ875860	784 bp	DNA	linear	GSS 08-NOV-1999
LOCUS	V130C9 mTn-3xHA/lacZ	Insertion Library, strain Y2278	Saccharomyces cerevisiae genomic 5', DNA sequence.			
DEFINITION	AQ875860					
ACCESSION	AQ875860.1	GI:6288104				
VERSION	GSS.					
KEYWORDS	baker's yeast.					
SOURCE	Saccharomyces cerevisiae					
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.					
REFERENCE	1 (bases 1 to 784)					
AUTHORS	Ross-Wadonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., deSteghes, S. A., Cheung, K. H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.					
TITLE	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Kumar A					
	Michael Snyder, Dept. of Mol. Cell. and Dev. Biology					
	Yale University					
	P.O. Box 208103, New Haven, CT 06520-8103, USA					
	Tel: 203 432 9949					
	Fax: 203 432 6161					
	Email: anuj.kumare@yale.edu					
	te of mTn-3xHA/lacZ insertion.					
	Seq primer: GGCCTCTCTTCTTTGGAAGTAC					

[illegible]

RESULT 7

A0872886

LOCUS

DEFINITION

ACCESSION

VERSION
REVISION

KEYWORDS

SOURCE

Db	259	GAGCCCAACGGGCTCCACTACCGCGCGGTCGCCCATTTTGGGAAGTCATCTCGTCCCAA	200
Qy	359	AAAGGAATAAGCCATAACATATCGTTACTGTTTTTGGAAACATCGCCGTTTTTCGCCCGATTTC	418
Db	199	AAAGGAATAAGCCATAACATATCGTTACTGTTTTTGGAAACATCGCCGTTTTTCGCCCGATTTC	140
Qy	419	CGCCTCAGCGGGTATAAAAAGAGATC	444
Db	139	CGCCTCAGCGGGTATAAAAAGAGATC	114
RESULT 9			
AQ501881/c			
LOCUS	AQ501881	466 bp	linear
DEFINITION	V13D1 mTn-3xHA/lacZ Insertion library		GSS 29-APR-1999
ACCSSION	genomic 5', DNA sequence.		
VERSION	AQ501881		
KEYWORDS	AQ501881.1	GI:4707531	
SOURCE	GSS.		
ORGANISM	baker's yeast.		
REFERENCE	Saccharomyces cerevisiae		
AUTHORS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
	1 (bases 1 to 466)		
	Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,		
	desEtages, S. A., Cheung, K. H., Sheehan, A., Symoniatidis, D., Jansen, R.,		
	Unasky, L., Heldtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,		
	Hager, K., Miller, P., Roeder, G. S., and Snyder, M.,		
	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and		
	Gene Disruption		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Kumar A		
	Michael Snyder, Dept. of Mol. Cell. and Dev. Biology		
	Yale University		
	P.O. Box 208103, New Haven, CT 06520-8103, USA		
	Tel: 203 432 9949		
	Fax: 203 432 6161		
	Email: anuj.kumar@yale.edu		
	te of mTn-3xHA/lacZ insertion.		
	Seq primer: GGCTTCCTTCCTTTGGAGTAC		
	Class: transposon-tagged.		
FEATURES	Location/Qualifiers		
source	1..466		
	/organism="Saccharomyces cerevisiae"		
	/db_xref="taxon:4932"		
	/clone_lib="mTn-3xHA/lacZ Insertion Library"		
	/lab_host="E. coli"		
	/note="Vector: pHS56-Sal; A yeast genomic DNA library		
	(lacking mitochondrial DNA) was prepared in pHS56-Sal;		
	genomic DNA was size-fractionated (DNA of roughly 2-3 kb		
	in length) prior to cloning. This library was		
	subsequently mutagenized with a mTn-3xHA/lacZ		
	mini-transposon containing lacZ, URA3, and tet resistance."		
BASE COUNT	131 a 92 c 120 g		
ORIGIN	119 t 4 others		
Query Match	31.7%;	Score 366.6;	DB 17; Length 466;
Best Local Similarity	94.6%;	Pred. No. 1.4e-70;	
Matches 389;	Conservative	0; Mismatches 21;	Indels 1; Gaps 1;
Qy	349	TCGCTCCCAAAAAGGAATAGCCATAACATATCGTTACTGTTTTTGGAAACATCGCCGTTTT	408
Db	466	TCGCTCCCAAAAAGGAATAGCCATAACATATCGTTACTGTTTTTGGAAACATCGCCGTTAT	408
Qy	409	CGCCCGAATCCGCGCTCAGCGGGTATAAAAAGAGATCTTTTTTTTTTCTGGCTGTCCTTC	468
Db	407	CGCCCGAATCCGCTCAGCGGGTATAAAAAGAGATCTTTTTTTTTTCTGGCTGTCCTTC	348
Qy	469	CATTTTAAATCTTATCTGCTCCCTTTGTGATCTTACGGTCTCACTAACCTCTCTTCAA	528
Db	347	CATTTTAAATCTTATCTGCTCCCTTTGTGATCTTACGGTCTCACTAACCTCTCTTCAA	288
Qy	529	CTGCTCAATAATTTCCCGCTATCGAAATTTCCCAAGACTACTTTTAAAGCTCAAAAATCGCT	588

```

Db      287  CTTGCTCAATAATTTCCCGCTATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGCT 228
      |||
QY      589  GCCAACACAAACAGCCCTTCCACATTCGCTACCGTGACCATGCGGGAATTTAGAAAGG 648
      |||
Db      227  GCCAACACAAACAGCCCTTCCACATTCGCTACCGTGACCATGCGGGAATTTAGAAAGG 168
      |||
QY      649  TGCCTTTGCCACCTATGCGTGAAGTTCCTATGTTCTTACTCAAAACTCCATGGGCAGCT 708
      |||
Db      167  TGCCTTTGCCACCTATGCGTGAAGTTCCTATGTTCTTACTCAAAACTCCATGGGCAGCT 108
      |||
QY      709  CCGCTTTCGCTCCGCTCTTCAATAGAAATGCGGAAAGGATTGGAGG 759
      |||
Db      107  CCGCTTTCGCTCCGCTCTTCAATAGAAATGCGGAAAGGATTGGAGG 57
      |||

RESULT 10
A0500866/c
LOCUS   V31A2 mtn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', DNA sequence.
ACCESSION A0500866
VERSION   A0500866.1 GI:4706516
KEYWORDS GSS.
SOURCE   baker's yeast.
ORGANISM Saccharomyces cerevisiae
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS 1 (bases 1 to 562)
          Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
          desEtages, S. A., Cheung, K. -H., Sheehan, A., Symoniatis, D., Jansen, R.,
          Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
          Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
          Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
          Gene Disruption
          Unpublished (1999)
          Contact: Kumar A
          Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
          Yale University
          P.O. Box 208103, New Haven, CT 06520-8103, USA
          Tel: 203 432 9949
          Fax: 203 432 6161
          Email: anuj.kumar@yale.edu
          te of mtn-3xHA/lacZ insertion.
          Seq primer: GGCCTTCTTCTTTGGAGATAC
          Class: transposon-tagged.
          Location/Qualifiers
            1..562
              /organism="Saccharomyces cerevisiae"
              /db_xref="taxon:4932"
              /clone_lib="mtn-3xHA/lacZ Insertion Library"
              /lab_host="E. coli"
              /note="Vector: pHS56-Sal; A yeast genomic DNA library
              (lacking mitochondrial DNA) was prepared in pHS56-Sal;
              genomic DNA was size-fractionated (DNA of roughly 2-3 kb
              in length) prior to cloning. This library was
              subsequently mutagenized with a mtn-3xHA/lacZ
              minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 195 a 87 c 127 g 153 t
ORIGIN

Query Match 30.0%; Score 347.4; DB 17; Length 562;
Best Local Similarity 96.8%; Pred. No. 2.2e-66;
Matches 365; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY      783  GACATGAACACAAACAAATTTGGTTCGGCGAACTAAATCTATCTTCAA-CCAGGGTAA 841
      |||
Db      561  GACATGTACACCAACAAATATGGTTCGGCGAACTAAATCTATGTCAACCGGGGTAA 502
      |||
QY      842  GGTGAGGAATGGACTTCTTAAGTTCCTTTTCATACCTCTTTCTTTCTTTCTTTCCATTTC 901
      |||
Db      501  GGTGAGGACATGGACTTCTTAAGTTCCTTTTCATACCTCTTTCTTTCTTTCTTTCCATTTC 442
      |||

```

```

QY      902  CCACTAGTCTGTTCTTTCTTTCTTTCTTTAGATACCCTTCTTTTCAGGACTCTCGTCTTAC 961
      |||
Db      441  CCACTAGTCTGTTCTTTCTTTCTTTCTTTAGATACCCTTCTTTTCAGGACTCTCGTCTTAC 382
      |||
QY      962  TATTCTTGTCAATCTTCGAAACATTTCTCTCCCGTGCATTTCTTTCCCTTTTATATACATA 1021
      |||
Db      381  TATTCTTGTCAATCTTCGAAACATTTCTCTCCCGTGCATTTCTTTCCCTTTTATATACATA 322
      |||
QY      1022 TATATATATATATATATATCTCTCTCTCTACGTATTTTGTATTTCTGTGCTTTATCA 1081
      |||
Db      321  TATATATATATATATATATCTCTCTCTCTACGTATTTTGTATTTCTGTGCTTTATCA 262
      |||
QY      1082 AGATAGTCTATAATACGTTTCATACAGCTAGATATPCGTAGCGCAACATTTGCCCTC 1141
      |||
Db      261  AGATAGTCTATAATACGTTTCATACAGCTAGATATPCGTAGCGCAACATTTGCCCTC 202
      |||
QY      1142 CTCCTTGATCAATGCTTT 1158
      |||
Db      201  CTCCTTGATCAATGCTTT 185
      |||

RESULT 11
A0500106/c
LOCUS   V30F8 mtn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', DNA sequence.
ACCESSION A0500106
VERSION   A0500106.1 GI:4704928
KEYWORDS GSS.
SOURCE   baker's yeast.
ORGANISM Saccharomyces cerevisiae
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS 1 (bases 1 to 543)
          Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
          desEtages, S. A., Cheung, K. -H., Sheehan, A., Symoniatis, D., Jansen, R.,
          Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
          Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
          Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
          Gene Disruption
          Unpublished (1999)
          Contact: Kumar A
          Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
          Yale University
          P.O. Box 208103, New Haven, CT 06520-8103, USA
          Tel: 203 432 9949
          Fax: 203 432 6161
          Email: anuj.kumar@yale.edu
          te of mtn-3xHA/lacZ insertion.
          Seq primer: GGCCTTCTTCTTTGGAGATAC
          Class: transposon-tagged.
          Location/Qualifiers
            1..543
              /organism="Saccharomyces cerevisiae"
              /db_xref="taxon:4932"
              /clone_lib="mtn-3xHA/lacZ Insertion Library"
              /lab_host="E. coli"
              /note="Vector: pHS56-Sal; A yeast genomic DNA library
              (lacking mitochondrial DNA) was prepared in pHS56-Sal;
              genomic DNA was size-fractionated (DNA of roughly 2-3 kb
              in length) prior to cloning. This library was
              subsequently mutagenized with a mtn-3xHA/lacZ
              minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 172 a 78 c 118 g 175 t
ORIGIN

Query Match 20.2%; Score 234.2; DB 17; Length 543;
Best Local Similarity 91.5%; Pred. No. 1.9e-41;
Matches 248; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY      888  TCTCTTCCATTTCCCACTAGTCTGTTCTTTCTTTCTTTCTTTAGATACCCTTTTTCAGG 947
      |||
Db      515  TCTCTTCCAAATCCCAAGATAAGTGGTATATTATATACCTAGATACCCTTTTTCAGG 456
      |||

```

[illegible][illegible]

```

source
l. 525
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/note="Vector: pHSS6-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS6-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mtn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT      157 a      89 c     111 g     168 t
ORIGIN
Query Match          11.8%;   Score 137.2;   DB 17;   Length 525;
Best Local Similarity 94.7%;   Pred. No. 41e-20;
Matches 142; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY    1009 CTTTATATACATATATATATATATATATATGTCTCTTCACGATTATTTGGTATTTC 1068
|||||
Db    525 CTTATATACATATATATATATATATATATATATATATGTCTCTTCACGAATTTTGGTATTTC 466
|||||
QY    1069 TGTGCTCTTTATCAAGATAGTCTTAATAACGTTTGATACAGCTTAGATATCGCTAGCGCCA 1128
|||||
Db    465 TGTGTTTTTTATCAAGATAGTCAATAAATACGTTTGATACAGATAGATATCGCTAGCGCCA 406
|||||

```


QY	1129	ACATTGTCCTCTCTTGATCAATGCTTT	1158
Db	405	ACATTGTCCTCTCTTGATCAATGCTTT	376
RESULT 14			
LOCUS	AQ501433/c		
DEFINITION	V22C5 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae genomic 5', DNA sequence.		
ACCESSION	AQ501433		
VERSION	AQ501433.1	GI:4707083	
KEYWORDS	GSS.		
SOURCE	baker's yeast.		
ORGANISM	Saccharomyces cerevisiae		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
AUTHORS	1 (bases 1 to 543) Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A., desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniat,D., Jansen,R., Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R., Hager,K., Miller,P., Roeder,G.S. and Snyder,M.		
TITLE	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Kumar A Michael Snyder, Dept. of Mol. Cell. and Dev. Biology Yale University P.O. Box 208103, New Haven, CT 06520-8103, USA Tel: 203 432 9949 Fax: 203 432 6161 Email: anuj.kumar@yale.edu te of mTn-3xHA/lacZ insertion. Seq primer: GGCCTCTTCTTGTGGAAGTAC Class: transposon-tagged.		
FEATURES	source		
	1..543		
	/organism="Saccharomyces cerevisiae"		
	/db_xref="taxon:4932"		
	/clone_lib="mTn-3xHA/lacZ Insertion Library"		
	/lab_host="E. coli"		
	/note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."		
BASE COUNT	128 a 151 c 127 g 136 t		
ORIGIN			
Query Match	6.6%; Score 76.4; DB 17; Length 543;		
Best Local Similarity	98.7%; Pred. No. 1.1e-06;		
Matches	77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	499	GATCTTACGGTCTCACTACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATT	558
Db	140	GATCTTACGGTCTCACTACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATT	81
QY	559	CCCAAGACTACTTTTACG 576	
Db	80	CCCAAGACTACTTTTACG 63	
RESULT 15			
LOCUS	AQ503366/c		
DEFINITION	V56E10 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae genomic 5', DNA sequence.		
ACCESSION	AQ503366		
VERSION	AQ503366.1	GI:4705912	
KEYWORDS	GSS.		
SOURCE	baker's yeast.		
ORGANISM	Saccharomyces cerevisiae		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
AUTHORS	1 (bases 1 to 530) Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A., desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniat,D., Jansen,R., Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R., Hager,K., Miller,P., Roeder,G.S. and Snyder,M.		
TITLE	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Kumar A Michael Snyder, Dept. of Mol. Cell. and Dev. Biology Yale University P.O. Box 208103, New Haven, CT 06520-8103, USA Tel: 203 432 9949 Fax: 203 432 6161 Email: anuj.kumar@yale.edu te of mTn-3xHA/lacZ insertion. Seq primer: GGCCTCTTCTTGTGGAAGTAC Class: transposon-tagged.		
FEATURES	source		
	1..530		
	/organism="Saccharomyces cerevisiae"		
	/db_xref="taxon:4932"		
	/clone_lib="mTn-3xHA/lacZ Insertion Library"		
	/lab_host="E. coli"		
	/note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."		
BASE COUNT	128 a 151 c 127 g 136 t		
ORIGIN			
Query Match	6.6%; Score 76.4; DB 17; Length 543;		
Best Local Similarity	98.7%; Pred. No. 1.1e-06;		
Matches	77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	499	GATCTTACGGTCTCACTACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATT	558
Db	140	GATCTTACGGTCTCACTACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATT	81
QY	559	CCCAAGACTACTTTTACG 576	
Db	80	CCCAAGACTACTTTTACG 63	
Search completed:	July 17, 2003, 13:18:10		
Job time :	1649 secs		

THIS PAGE BLANK (USP10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:51:17 ; Search time 14 Seconds
(without alignments)
218.570 Million cell updates/sec

Title: US-09-814-661a-2
Perfect score: 104
Sequence: 1 MONSDYFYAQNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size: 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	7.7	523	3	US-08-606-505B-67
2	8	7.7	523	4	US-09-616-990-67
3	8	7.7	651	1	US-08-431-080-24
4	8	7.7	651	2	US-08-938-534-24
5	8	7.7	651	4	US-09-345-294-24
6	8	7.7	738	3	US-08-864-038A-3
7	8	7.7	909	4	US-09-425-383-2
8	7	6.7	15	4	US-08-218-369-1
9	7	6.7	15	5	PCT-US95-03742-1
10	7	6.7	54	1	US-08-605-053-1
11	7	6.7	539	2	US-08-978-182-3
12	7	6.7	539	2	US-09-205-681-3
13	7	6.7	606	4	US-09-382-123A-6
14	7	6.7	664	3	US-09-295-186-17
15	7	6.7	694	3	US-08-559-397A-31
16	7	6.7	1381	4	US-09-540-245A-16
17	6	5.8	7	2	US-08-769-745-7
18	6	5.8	18	4	US-08-944-604-24
19	6	5.8	21	2	US-08-530-326-8
20	6	5.8	21	4	US-08-530-342A-8
21	6	5.8	26	2	US-08-479-614-11
22	6	5.8	44	1	US-08-479-577-2
23	6	5.8	44	2	US-08-756-977-2
24	6	5.8	44	4	US-09-115-446-8
25	6	5.8	44	5	PCT-US93-05331-6
26	6	5.8	73	2	US-08-248-839C-114
27	6	5.8	108	2	US-08-825-782-1

28	6	5.8	108	2	US-08-825-782-3
29	6	5.8	108	2	US-08-825-782-4
30	6	5.8	124	1	US-08-276-852-146
31	6	5.8	124	1	US-08-899-575-146
32	6	5.8	124	1	US-08-899-575-146
33	6	5.8	124	5	PCT-US95-08743-146
34	6	5.8	142	4	US-09-370-838-212
35	6	5.8	151	4	US-09-228-986-94
36	6	5.8	154	2	US-08-479-614-19
37	6	5.8	173	2	US-08-658-639-12
38	6	5.8	173	4	US-08-944-604-12
39	6	5.8	176	2	US-08-989-286-3
40	6	5.8	178	2	US-08-630-822A-100
41	6	5.8	178	2	US-09-005-069-100
42	6	5.8	178	4	US-09-171-156A-49
43	6	5.8	222	4	US-08-944-604-16
44	6	5.8	223	4	US-09-009-816-4
45	6	5.8	223	6	5218099-5

ALIGNMENTS

RESULT 1
US-08-606-505B-67
; Sequence 67, Application US/08606505B
; Patent No. 6114601
; GENERAL INFORMATION:
; APPLICANT: KIKUCHI, Yasuhiro
; APPLICANT: KIKUCHI, Yasuhiro
; APPLICANT: SHIMADA, Yukihisa
; APPLICANT: OHBAYASHI, Masaya
; APPLICANT: SHIMADA, Ritsuko
; APPLICANT: OKINAKA, Yasushi
; TITLE OF INVENTION: NOVEL PLANT GENES
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO
; STREET: 30. Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112-3801
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 720 Kb storage.
; OPERATING SYSTEM: MS-DOS Ver3.30
; SOFTWARE: PATENT AID Ver1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/606,505B
; FILING DATE: 23-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP44963/92
; FILING DATE: 02-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Perry, Lawrence S.
; REGISTRATION NUMBER: 31865
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-218-2100
; TELEFAX: 212-218-2200
; INFORMATION FOR SEQ ID NO: 67 :
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Campanula medium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 180 to 1748
; IDENTIFICATION METHOD: by experiment

Sequence 3, Appli
Sequence 4, Appli
Sequence 146, App
Sequence 146, App
Sequence 146, App
Sequence 146, App
Sequence 94, Appl
Sequence 212, App
Sequence 19, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 3, Appli
Sequence 100, App
Sequence 100, App
Sequence 49, Appl
Sequence 16, Appl
Sequence 4, Appli
Patent No. 5218099

US-08-606-505B-67

Query Match 7.7%; Score 8; DB 3; Length 523;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
IIIIIIII
DB 209 SSASASAS 216

RESULT 2

US-09-616-990-67
Sequence 67, Application US/09616990
Patent No. 6232109
GENERAL INFORMATION:

APPLICANT: KIKUCHI, Yasuhiro
KIYOKAWA, Shigeto
SHIMADA, Yukihisa
OHYAYASHI, Masaya
SHIMADA, Ritsuko
OKINAKA, Yasushi

TITLE OF INVENTION: NOVEL PLANT GENES
NUMBER OF SEQUENCES: 67

CORRESPONDENCE ADDRESS:
ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112-3801

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.50 inch, 720 Kb storage.
COMPUTER: IBM PS/V
OPERATING SYSTEM: MS-DOS Ver3.30
SOFTWARE: PATENT AID Ver1.0
CURRENT APPLICATION DATA:
FILING DATE: 14-Jul-2000
APPLICATION NUMBER: US/09/616,990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP44963/92
FILING DATE: 02-MAR-1992

ATTORNEY/AGENT INFORMATION:
NAME: Perry, Lawrence S.
REGISTRATION NUMBER: 31865

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-218-2100
TELEFAX: 212-218-2200

INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 523 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Campanula medium

FEATURE:
NAME/KEY: CDS
LOCATION: 180 to 1748
IDENTIFICATION METHOD: by experiment

SEQUENCE DESCRIPTION: SEQ ID NO: 67

US-09-616-990-67
Query Match 7.7%; Score 8; DB 4; Length 523;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
IIIIIIII
DB 209 SSASASAS 216

RESULT 3

US-08-431-080-24
Sequence 24, Application US/08431080
Patent No. 5698686
GENERAL INFORMATION:

APPLICANT: Gottschling, Daniel E.
APPLICANT: Singer, Miriam S.
TITLE OF INVENTION: Telomerase Compositions and Methods
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TEXAS
COUNTRY: UNITED STATES OF AMERICA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,080
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER: SN 08/326,781
FILING DATE: October 20, 1994
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: ARCD:155/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924

INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 651 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-431-080-24
Query Match 7.7%; Score 8; DB 1; Length 651;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
IIIIIIII
DB 39 SASASASS 46

RESULT 4

US-08-938-534-24
Sequence 24, Application US/08938534
Patent No. 5916752
GENERAL INFORMATION:

APPLICANT: Gottschling, Daniel E.
APPLICANT: Singer, Miriam S.
TITLE OF INVENTION: Telomerase Compositions and Methods
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TEXAS
COUNTRY: UNITED STATES OF AMERICA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,534
FILING DATE: 26-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431,080
FILING DATE:
APPLICATION NUMBER: SN 08/326,781
FILING DATE: October 20, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:155/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 651 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-938-534-24

Query Match 7.7%; Score 8; DB 2; Length 651;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
DB 39 SASASASS 46

RESULT 5
US-09-345-294-24
Sequence 24, Application US/09345294
Patent No. 6387619
GENERAL INFORMATION:
APPLICANT: Gottschling, Daniel E.
TITLE OF INVENTION: Telomerase Compositions and Methods
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TEXAS
COUNTRY: UNITED STATES OF AMERICA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/345,294
FILING DATE: 30-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431,080
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:155/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:
LENGTH: 651 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-345-294-24

Query Match 7.7%; Score 8; DB 4; Length 651;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
DB 39 SASASASS 46

RESULT 6
US-08-864-038A-3
Sequence 3, Application US/08864038A
Patent No. 6001592
GENERAL INFORMATION:
APPLICANT: Kunio NAKASHIMA et al.
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
TITLE OF INVENTION: TO SAID POLYPEPTIDE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: 812-5 Hirano
STREET: Isshinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 7.7%; Score 8; DB 3; Length 738;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
|||||
Db 422 SSASASAS 429

RESULT 7

US-09-425-383-2

; Sequence 2, Application US/09425383

; Patent No. 6194637

; GENERAL INFORMATION:

; APPLICANT: Mahajan, Pramod B.

; APPLICANT: Shi, Jinrui

; TITLE OF INVENTION: Malze DNA Ligase I Orthologue and Uses

; TITLE OF INVENTION: Thereof

; FILE REFERENCE: 0962

; CURRENT APPLICATION NUMBER: US/09/425,383

; CURRENT FILING DATE: 1999-10-22

; EARLIER APPLICATION NUMBER: 60/108,793

; EARLIER FILING DATE: 1998-11-17

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO. 2

; LENGTH: 909

; TYPE: PRT

; ORGANISM: Zea mays

US-09-425-383-2

Query Match 7.7%; Score 8; DB 4; Length 909;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61

|||||

Db 49 SASASASS 56

RESULT 8

US-08-218-369-1

; Sequence 1, Application US/08218369

; Patent No. 6312699

; GENERAL INFORMATION:

; APPLICANT: Curjel, David T.

; APPLICANT: Engler, Jeffrey A.

; TITLE OF INVENTION: Ligands Added to Adenovirus Fiber

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Patrea L. Pabst

; STREET: 1100 Peachtree Street, Suite 2800

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30309-4530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/218,369

; FILING DATE: 28-MAR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Pabst, Patrea L.

; REGISTRATION NUMBER: 31,284

; REFERENCE/DOCKET NUMBER: IGI101

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (404) 815-6508

; TELEFAX: (404) 815-6555

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; FRAGMENT TYPE: Internal
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..15
; OTHER INFORMATION: /note= "Linker peptide attached at the carboxyl end of the
US-08-218-369-1

Query Match 6.7%; Score 7; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASAS 60

|||||

Db 5 SASASAS 11

RESULT 9

PCT-US95-03742-1

; Sequence 1, Application PC/TUS9503742

; GENERAL INFORMATION:

; APPLICANT: The UAB Research Foundation

; TITLE OF INVENTION: Ligands Added to Adenovirus Fiber

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Patrea L. Pabst

; STREET: 2800 One Atlantic Center

; STREET: 1201 West Peachtree Street

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30309-3450

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/03742

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Pabst, Patrea L.

; REGISTRATION NUMBER: 31,284

; REFERENCE/DOCKET NUMBER: IGI101

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (404) 873-8794

; TELEFAX: (404) 873-8795

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: YES

; FRAGMENT TYPE: Internal

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..15

; OTHER INFORMATION: /note= "Linker peptide attached

; OTHER INFORMATION: at the carboxyl end of the fiber protein."

PCT-US95-03742-1

Query Match 6.7%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASAS 60

|||||

Db 5 SASASAS 11

RESULT 10
 US-08-605-053-1
 ; Sequence 1, Application US/08605053
 ; Patent No. 5741674
 ; GENERAL INFORMATION:
 ; APPLICANT: SCHWEDEN, Juergen
 ; APPLICANT: BOLLSCHWEILER, Claus
 ; APPLICANT: PIONTEK, Michael
 ; APPLICANT: WEYDEMANN, Ulrike
 ; APPLICANT: JANOWICZ, Zbigniew A.
 ; APPLICANT: STRASSER, Alexander W.M.
 ; TITLE OF INVENTION: THE RECOMBINANT PRODUCTION OF PROTEINS IN
 ; TITLE OF INVENTION: YEAST
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Keil & Weinkauff
 ; STREET: 1101 Connecticut Avenue
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
 ; COMPUTER: IBM compatible
 ; OPERATING SYSTEM: MS-DOS version 6.0
 ; SOFTWARE: WordPerfect version 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/605.053
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/EP94/02897
 ; FILING DATE: 01-SEP-1994
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 54 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE: Schwanniomycetes occidentalis
 ; US-08-605-053-1

Query Match 6.7%; Score 7; DB 1; Length 54;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 GSSASAS 58
 Db 7 GSSASAS 13

RESULT 11
 US-08-978-182-3
 ; Sequence 3, Application US/08978182
 ; Patent No. 5849556
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Kaser, Matthew
 ; APPLICANT: Mathur, Preete
 ; TITLE OF INVENTION: HUMAN GROWTH-RELATED CDC10 HOMOLOG
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/978,182
 FILING DATE: Herewith
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0426 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555
 TELEFAX: 650-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 539 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 508229
 US-08-978-182-3

Query Match 6.7%; Score 7; DB 2; Length 539;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 ASASASS 61
 Db 36 ASASASS 42

RESULT 12
 US-09-205-681-3
 ; Sequence 3, Application US/09205681
 ; Patent No. 5952214
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Kaser, Matthew
 ; APPLICANT: Mathur, Preete
 ; TITLE OF INVENTION: HUMAN GROWTH-RELATED CDC10 HOMOLOG
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA: US/09/205,681
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/978,182
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PE-0426 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-853-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 508229
US-09-205-681-3

Query Match 6.7%; Score 7; DB 2; Length 539;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 ASASASS 61
Db 36 ASASASS 42

RESULT 13
US-09-362-123A-6
Sequence 6, Application US/09362123A
Patent No. 6451558
GENERAL INFORMATION:
APPLICANT: Cooke, Michael Paul
APPLICANT: Holness, Claire
APPLICANT: Sirenko, Oksana
TITLE OF INVENTION: No. 6451558el Genes in the Control of Hematopoiesis
FILE REFERENCE: 4-30629A/SYS
CURRENT APPLICATION NUMBER: US/09/362.123A
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: 09/128,310
PRIOR FILING DATE: 1998-08-03
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 606
TYPE: PRT
ORGANISM: Homo sapiens
US-09-362-123A-6

Query Match 6.7%; Score 7; DB 4; Length 606;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASAS 60
Db 558 SASASAS 564

RESULT 14
US-09-295-186-17
Sequence 17, Application US/09295186B
Patent No. 6127137
GENERAL INFORMATION:
APPLICANT: Hasida, Miyoko
APPLICANT: Tsutsumi, No. 6127137iko
APPLICANT: Halkier, Torben
APPLICANT: Stringer, Mary Ann
TITLE OF INVENTION: An Acidic Phospholipase, Production, and
METHODS OF USING THEREOF (As Amended)
FILE REFERENCE: 4953.204-US
CURRENT APPLICATION NUMBER: US/09/295,186B
PRIOR FILING DATE: 1999-04-20
PRIOR APPLICATION NUMBER: 1215/96
PRIOR FILING DATE: 1996-10-31

PRIOR APPLICATION NUMBER: PCT/DK97/00490
PRIOR FILING DATE: 1997-10-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 664
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-295-186-17

Query Match 6.7%; Score 7; DB 3; Length 664;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASAS 60
Db 612 SASASAS 618

RESULT 15
US-08-559-397A-31
Sequence 31, Application US/08559397A
Patent No. 6083713
GENERAL INFORMATION:
APPLICANT: Manly, Susan P.
APPLICANT: Kozlowski, Michael R.
APPLICANT: Neve, Rachael L.
TITLE OF INVENTION: CLONING AND EXPRESSION OF
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,397A
FILING DATE: 15-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6013-135
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-559-397A-31

Query Match 6.7%; Score 7; DB 3; Length 694;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASA 59
Db 261 SSASASA 267

Search completed: July 16, 2003, 12:53:32

Job time : 15 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 20:24:47 ; Search time 172 Seconds
(without alignments)
1247.396 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 104

Sequence: 1 MQNQDYFYAQRCCQQAP.....FGSGELKSMFNQCKVEEMDF 104

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2875925

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09814661/runat_16072003_115403_15732/app_query.fasta_1.263
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09814661@cgn_1.1.80 @runat_16072003_115403_15732
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	104	100.0	1158	11 US-09-814-661A-1

c	2	65	11	US-09-814-661A-6
	3	112	11	US-09-861-687-10
	4	222	10	US-09-923-876-116
	5	8.7	450	US-09-934-455-265
c	6	800	10	US-09-910-943-585
c	7	1158	15	US-10-156-761-6806
	8	2263	15	US-10-027-632-266262
	9	8.7	2263	US-10-027-632-266263
c	10	9025608	15	US-10-156-761-1
	11	60	15	US-10-156-824A-26
c	12	171	11	US-09-878-574-7681
c	13	259	11	US-09-878-574-5555
c	14	289	10	US-09-728-446-1452
c	15	306	10	US-09-923-876-3699
c	16	372	10	US-09-815-242-6975
c	17	409	11	US-09-960-352-4093
	18	423	15	US-10-156-761-3496
c	19	431	11	US-09-974-300-3487
c	20	468	11	US-09-894-844-29
c	21	488	12	US-09-918-995-2496
c	22	502	12	US-09-918-995-31644
c	23	528	11	US-09-974-300-1583
c	24	532	15	US-10-027-632-275343
c	25	586	15	US-10-198-846-8816
c	26	600	10	US-09-736-611-5
c	27	600	10	US-09-736-611-7
c	28	600	10	US-09-740-359-4
c	29	600	10	US-09-740-359-6
c	30	600	11	US-09-894-711-4
c	31	600	11	US-09-894-711-6
c	32	629	15	US-10-178-213-46
c	33	631	15	US-10-178-213-55
c	34	639	15	US-10-198-846-9481
c	35	675	15	US-10-198-846-2791
c	36	700	10	US-09-791-171-53
c	37	739	15	US-10-027-632-158839
c	38	756	11	US-09-962-832-255
c	39	756	11	US-09-999-699-5
c	40	762	11	US-09-974-300-6814
c	41	816	11	US-09-738-626-38
c	42	957	15	US-10-156-761-5024
c	43	1020	11	US-09-738-626-1733
c	44	1025	10	US-09-770-445-200
c	45	1029	15	US-10-156-761-5681

ALIGNMENTS

RESULT 1
US-09-814-661A-1
; Sequence 1, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; FILE REFERENCE: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814, 661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: S. Cerevisiae
US-09-814-661A-1

Alignment Scores:
Pred. No.: 5.58e-98 Length: 1158
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-814-661A-1 (1-1158)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
DB 549 ATGCAAAATCCCAAGACTACTTTTACGCTCAAAATCGCTGCCAACAAACAGCCCT 608
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuProMetAla 40
DB 609 TCCACATTCGTCACCGTACCATGGGAAATTTAGAAAGGTCCTTTGCCACCTATGGCT 668
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
DB 669 GAGGTTCCTATGCTCTACTCAAACTCCATGGGAGCTCGCTTCCTCCGCTTCT 728
QY 61 SerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAspHisAspMet 80
DB 729 TCATTAGAAATGCGGAAAGGATTTGGAGGAGAGACTCAACTCTATCATCATGACATG 788
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
DB 789 ACAACAACAATAATTTGGTTCTGCGGAATAAAATCTATGTTCAACGAGGTAAGGTCGAG 848
QY 101 GluMetAspPhe 104
DB 849 GAAATGGACTTC 860

RESULT 2

US-09-814-661A-6/c
; Sequence 6, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE SU
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814,661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer
US-09-814-661A-6

Alignment Scores:
Pred. No.: 0.369 Length: 65
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 11 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-814-661A-6 (1-65)

QY 96 GlnGlyLysValGluGluMetAspPhe 104
DB 44 CAGGGTAAGGTCGAGGAAATGGACTTC 18

RESULT 3

US-09-861-687-10
; Sequence 10, Application US/09861687
; Publication No. US20020193292A1
; GENERAL INFORMATION:
; APPLICANT: Markussen, Jan
; Jonassen, Ib
; Havelund, Svend

Brandt, Jakob
Kurtzhals, Peter
Hansen, Hertz Per
Kaarsholm, Niels Christibn
TITLE OF INVENTION: INSULIN DERIVATIVES
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20020193292A1b No. US20020193292A1disk of No. US200201
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/861,687
FILING DATE: 21-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/932,082
FILING DATE: 16-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4341.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 112 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-861-687-10

Alignment Scores:
Pred. No.: 0.595 Length: 112
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 11 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-861-687-10 (1-112)

QY 54 SerAlaSerAlaSerAlaSerSerLeu 62
DB 69 TCAGCTTCAGCTTCAGCTTCCTCTCT 95

RESULT 4

US-09-923-876-116
; Sequence 116, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalugudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program

; SEQ ID NO 116
; LENGTH: 222
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700142585H1
; NAME/KEY: unsure
; LOCATION: 195
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-116

Alignment Scores:
Pred. No.: 1.09 Length: 222
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 10 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-923-876-116 (1-222)

QY 54 SerAlaSerAlaSerAlaSerSerLeu 62
DB 4 AGTCCAGTGCAGCGCGTCTCACTT 30

RESULT 5

US-09-934-455-265
; Sequence 265, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaira
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934,455
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 265
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27)..(422)
; OTHER INFORMATION: G1795
US-09-934-455-265

Alignment Scores:
Pred. No.: 2.02 Length: 450
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 12 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-934-455-265 (1-450)
QY 53 SerSerAlaSerAlaSerAlaSerSer 61
DB 297 TCCTCCGCTCCGCTCGCTTCTTCT 323

RESULT 6

US-09-910-943-585/C
; Sequence 585, Application US/09910943
; Patent No. US20020081610A1
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Brivanlou, Ali
; APPLICANT: Altman, Curtis
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
; FILE REFERENCE: 7529/1G148US1
; CURRENT APPLICATION NUMBER: US/09/910,943
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 742
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 585
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(800)
; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-585

Alignment Scores:
Pred. No.: 3.35 Length: 800
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 10 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-910-943-585 (1-800)

QY 53 SerSerAlaSerAlaSerAlaSerSer 61
DB 701 TCATCAGCATCAGCTCGCATCTTCA 675

RESULT 7

US-10-156-761-6806/C
; Sequence 6806, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6806
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1158)
US-10-156-761-6806

Alignment Scores:
Pred. No.: 4.64 Length: 1158
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 15 Gaps: 0

US-09-814-661A-2 (1-104) x US-10-156-761-6806 (1-1158)

Qy 54 SerAlaSerAlaSerAlaSerLeu 62
Db 959 TCGCCAGCGCTGCTGCTCTCTGTTG 933

RESULT 8
US-10-027-632-266262
; Sequence 266262, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 266262
; LENGTH: 2263
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-266262

Alignment Scores:
Pred. No.: 8.37 Length: 2263
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 15 Gaps: 0

US-09-814-661A-2 (1-104) x US-10-027-632-266262 (1-2263)

Qy 54 SerAlaSerAlaSerAlaSerLeu 62
Db 257 TCTGCCAGTGCCTCTGCTTCTCTACTG 283

RESULT 9
US-10-027-632-266263
; Sequence 266263, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 266263
; LENGTH: 2263
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-266263

Alignment Scores:
Pred. No.: 8.37 Length: 2263
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 15 Gaps: 0

US-09-814-661A-2 (1-104) x US-10-027-632-266263 (1-2263)

Qy 54 SerAlaSerAlaSerAlaSerLeu 62
Db 257 TCTGCCAGTGCCTCTGCTTCTCTACTG 283

RESULT 10
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.23e+04 Length: 9025608
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 15 Gaps: 0

US-09-814-661A-2 (1-104) x US-10-156-761-1 (1-9025608)

QY 54 SerAlaSerAlaSerAlaSerLeu 62
 Db 8152802 TCGGCAGCGCGCTCGCGTCTCGTG 8152776

RESULT 11

US-10-156-824A-26
 ; Sequence 26, Application US/10156824A
 ; Publication No. US20030108521A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Calatrava, Manuel Rosa
 ; TITLE OF INVENTION: Adenovirus Protein IX, its Domain Involved in Capsid Assembly,
 ; FILE REFERENCE: 032751-065
 ; CURRENT APPLICATION NUMBER: US/10/156,824A
 ; CURRENT FILING DATE: 2002-11-14
 ; PRIOR APPLICATION NUMBER: US 60/293,974
 ; PRIOR FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 26
 ; LENGTH: 60
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: mutagenesis oligonucleotide
 US-10-156-824A-26

Alignment Scores:
 Pred. No.: 3.69 Length: 60
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.69% Indels: 0
 DB: 15 Gaps: 0

US-09-814-661A-2 (1-104) x US-10-156-824A-26 (1-60)

QY 53 SerSerAlaSerAlaSerAlaSer 60
 Db 3 TCCAGCGCGCTCGCGCTCGCGTCTCG 26

RESULT 12

US-09-878-574-7681/C
 ; Sequence 7681, Application US/09878574
 ; Patent No. US20020110548A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Byrum, Joseph R.
 ; APPLICANT: Thompson, Michael D.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(15401)B
 ; CURRENT APPLICATION NUMBER: US/09/878,574
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 09/333,535
 ; PRIOR FILING DATE: 1999-06-14
 ; NUMBER OF SEQ ID NOS: 15775
 ; SEQ ID NO 7681
 ; LENGTH: 171
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; OTHER INFORMATION: Clone ID: 701100205H1
 US-09-878-574-7681

Alignment Scores:
 Pred. No.: 9.27 Length: 171
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.69% Indels: 0
 DB: 11 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-878-574-7681 (1-171)

QY 54 SerAlaSerAlaSerAlaSerSer 61
 Db 94 TCAGCATCAGCCTCGCGCTCTCTCC 71

RESULT 13

US-09-878-574-5555/C
 ; Sequence 5555, Application US/09878574
 ; Patent No. US20020110548A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Byrum, Joseph R.
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Thompson, Michael D.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(15401)B
 ; CURRENT APPLICATION NUMBER: US/09/878,574
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 09/333,535
 ; PRIOR FILING DATE: 1999-06-14
 ; NUMBER OF SEQ ID NOS: 15775
 ; SEQ ID NO 5555
 ; LENGTH: 259
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; OTHER INFORMATION: Clone ID: 701097109H1
 US-09-878-574-5555

Alignment Scores:
 Pred. No.: 13.4 Length: 259
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.69% Indels: 0
 DB: 11 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-878-574-5555 (1-259)

QY 53 SerSerAlaSerAlaSerAlaSer 60
 Db 227 AGCTCAGCATCAGCATCAGCATCA 204

RESULT 14

US-09-728-446-1452/C
 ; Sequence 1452, Application US/09728446
 ; Patent No. US20020081668A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Friedrich, Glenn
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; TITLE OF INVENTION: No. US20020081668A1 Murine Polynucleotide Sequences
 ; FILE REFERENCE: LEX-0101-USA
 ; CURRENT APPLICATION NUMBER: US/09/728,446
 ; CURRENT FILING DATE: 2000-11-30
 ; PRIOR APPLICATION NUMBER: US 60/168,270
 ; PRIOR FILING DATE: 1999-12-01
 ; NUMBER OF SEQ ID NOS: 1461
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1452
 ; LENGTH: 289
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(289)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-728-446-1452

Alignment Scores:
 Pred. No.: 14.7 Length: 289
 Score: 8.00 Matches: 8

Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 7.69%
 DB: 10
 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-728-446-1452 (1-289)

Qy 53 SerSerAlaSerAlaSer 60
 Db 115 TCTCTGCTTCNGCTTCGCTTCT 92

RESULT 15

US-09-923-876-3699/c
 ; Sequence 3699, Application US/09923876
 ; Patent No. US20020013958A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath V.
 ; APPLICANT: Kamigaki, Laura Y. (Ito)
 ; APPLICANT: Sherman, Bradley K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
 ; FILE REFERENCE: PL-0012-1 CON
 ; CURRENT APPLICATION NUMBER: US/09/923,876
 ; CURRENT FILING DATE: 2001-08-06
 ; PRIOR APPLICATION NUMBER: 09/298,329
 ; PRIOR FILING DATE: 1999-04-21
 ; PRIOR APPLICATION NUMBER: 60/085,331
 ; PRIOR FILING DATE: 1998-05-05
 ; NUMBER OF SEQ ID NOS: 6332
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 3699
 ; LENGTH: 306
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; OTHER INFORMATION: Incyte ID No. US20020013958A1 700405451H1
 US-09-923-876-3699

Alignment Scores:

Pred. No.: 15.5
 Score: 8.00
 Length: 306
 Matches: 8
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 7.69%
 DB: 10
 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-923-876-3699 (1-306)

Qy 54 SerAlaSerAlaSerAlaSer 61
 Db 144 TCAGCTTCAGCTTCGCTTCATCT 121

Search completed: July 17, 2003, 22:12:16
 Job time : 1981 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 19:48:10 ; Search time 65 Seconds
(without alignments)
490.683 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 104
Sequence: 1 MNSQDYFYAQRCCQQOAP.....FGSGELKSMFNQGVKVEEMDF 104

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 1533381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09814661/runat_16072003_115401_15647/app_query.fasta_1.263
-DB=Issued_Patents_NA -QFMT=fascap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn.1.1.61@runat_16072003_115401_15647 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	8.7	112	4	US-08-932-082-10
c 2	9	8.7	219	1	US-08-468-674B-70
c 3	9	8.7	219	1	US-08-780-571-70
c 4	9	8.7	379	1	US-08-468-674B-74
c 5	9	8.7	379	1	US-08-780-571-74
c 6	9	8.7	538	1	US-08-400-256-47
c 7	9	8.7	538	3	US-08-975-365-47
c 8	8	7.7	63	1	US-08-400-256-12
c 9	8	7.7	63	3	US-08-975-365-12
c 10	8	7.7	72	4	US-09-097-767A-8
c 11	8	7.7	79	4	US-09-097-767A-7
c 12	8	7.7	82	1	US-08-468-674B-64

Sequence 64, Appl	82	7.7	1	US-08-780-571-64
Sequence 2, Appl	471	7.7	2	US-08-883-070-2
Sequence 36, Appl	533	7.7	4	US-09-097-767A-36
Sequence 44, Appl	535	7.7	1	US-08-400-256-44
Sequence 44, Appl	535	7.7	3	US-08-975-365-44
Sequence 2, Appl	867	7.7	2	US-08-961-858-2
Sequence 4, Appl	867	7.7	2	US-08-961-858-4
Sequence 2, Appl	867	7.7	3	US-09-089-593-2
Sequence 4, Appl	867	7.7	3	US-09-089-593-4
Sequence 3, Appl	867	7.7	3	US-08-950-925-3
Sequence 16, Appl	944	7.7	4	US-09-097-767A-16
Sequence 20, Appl	1021	7.7	4	US-08-818-111-20
Sequence 20, Appl	1021	7.7	4	US-08-818-111-20
Sequence 20, Appl	1021	7.7	4	US-09-056-556-20
Sequence 20, Appl	1021	7.7	4	US-09-072-596-20
Sequence 13, Appl	1053	7.7	4	US-09-097-767A-13
Sequence 9, Appl	1059	7.7	4	US-09-097-767A-9
Sequence 22, Appl	1244	7.7	4	US-09-097-767A-22
Sequence 7, Appl	1245	7.7	4	US-09-097-767A-25
Sequence 7, Appl	1260	7.7	4	US-09-387-212-7
Sequence 1, Appl	1260	7.7	4	US-09-948-802-7
Sequence 1, Appl	1569	7.7	4	US-09-368-588-1
Sequence 64, Appl	1927	7.7	3	US-08-606-505B-64
Sequence 64, Appl	1927	7.7	4	US-09-616-990-64
Sequence 23, Appl	2117	7.7	1	US-08-431-080-23
Sequence 23, Appl	2117	7.7	2	US-08-938-534-23
Sequence 23, Appl	2117	7.7	4	US-09-345-294-23
Sequence 1, Appl	2214	7.7	3	US-08-864-038A-1
Sequence 17, Appl	2403	7.7	1	US-08-441-139-17
Sequence 1, Appl	2875	7.7	2	US-08-883-070-1
Sequence 1, Appl	2941	7.7	4	US-09-425-383-1
Sequence 2, Appl	3331	7.7	3	US-08-864-038A-2
Sequence 4, Appl	3331	7.7	3	US-08-864-038A-4

ALIGNMENTS

RESULT 1

US-08-932-082-10
; Sequence 10, Application US/08932082
; Patent No. 6251856
; GENERAL INFORMATION:
; APPLICANT: Markussen, Jan
; APPLICANT: Jonassen, Ib
; APPLICANT: Havelund, Svend
; APPLICANT: Brandt, Jakob
; APPLICANT: Kurtzhals, Peter
; APPLICANT: Hansen, Hertz Per
; APPLICANT: Kaarsholm, Niels Christian
; TITLE OF INVENTION: INSULIN DERIVATIVES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 62518560 No. 6251856disk of No. 6251856th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,082
; FILING DATE: 12-AUG-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4341.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 112 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-932-082-10

Alignment Scores:
Pred. No.: 0.752 Length: 112
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 4 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-932-082-10 (1-112)

QY 54 SerAlaSerAlaSerAlaSerLeu 62
DB 69 TCAGCTTCAGCTTCAGCTTCCTCTT 95

RESULT 2

US-08-468-674B-70/c
Sequence 70, Application US/08468674B
Patent No. 5639642
GENERAL INFORMATION:
APPLICANT: Kjeldsen, Thomas B
APPLICANT: Vad, Knud
TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
NUMBER OF SEQUENCES: 89
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,674B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/282,852
FILING DATE: 29-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4085.220-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..198
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..219
OTHER INFORMATION: /note= "complementary strand"

US-08-468-674B-70

Alignment Scores:
Pred. No.: 1.37 Length: 219
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-468-674B-70 (1-219)

QY 54 SerAlaSerAlaSerAlaSerLeu 62
DB 29 TCAGCTTCAGCTTCAGCTTCCTCTC 3

RESULT 3

US-08-780-571-70/c
Sequence 70, Application US/08780571
Patent No. 5795746
GENERAL INFORMATION:
APPLICANT: Kjeldsen, Thomas B
APPLICANT: Vad, Knud
TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
NUMBER OF SEQUENCES: 89
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5795746o No. 5795746disk of No. 5795746th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/780,571
FILING DATE: 08-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/468,674
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: US 08/282,852
FILING DATE: 29-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4085.220-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..198
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..219
OTHER INFORMATION: /note= "complementary strand"

US-08-780-571-70

Alignment Scores:
Pred. No.: 1.37 Length: 219
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 1 Gaps: 0

```
US-09-814-661A-2 (1-104) x US-08-780-571-70 (1-219)
QY 54 SerAlaSerAlaSerAlaSerLeu 62
Db 29 TCAGCTTCAGCTTCAGCTTCTCTC 3
RESULT 4
US-08-468-674B-74/c
; Sequence 74, Application US/08468674B
; Patent No. 5639642
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642disk of No. 5639642th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,674B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/282,852
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4085.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 113..376
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..379
; OTHER INFORMATION: /note= "complementary strand"
US-08-468-674B-74
Alignment Scores:
Pred. No.: 2.25 Length: 379
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 1 Gaps: 0
US-09-814-661A-2 (1-104) x US-08-468-674B-74 (1-379)
QY 54 SerAlaSerAlaSerAlaSerLeu 62
Db 321 TCAGCTTCAGCTTCAGCTTCTCTC 295
RESULT 5
US-08-780-571-74/c
; Sequence 47, Application US/08400256
; Patent No. 5750497
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asser Sloth
; Sequence 74, Application US/08780571
; Patent No. 5795746
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5795746o No. 5795746disk of No. 5795746th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/780,571
; FILING DATE: 08-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/468,674
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/282,852
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4085.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 113..376
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..379
; OTHER INFORMATION: /note= "complementary strand"
US-08-780-571-74
Alignment Scores:
Pred. No.: 2.25 Length: 379
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 1 Gaps: 0
US-09-814-661A-2 (1-104) x US-08-780-571-74 (1-379)
QY 54 SerAlaSerAlaSerAlaSerLeu 62
Db 321 TCAGCTTCAGCTTCAGCTTCTCTC 295
RESULT 6
US-08-400-256-47/c
; Sequence 47, Application US/08400256
; Patent No. 5750497
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asser Sloth
```

APPLICANT: Markussen, Jan
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985.220-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 77..514
US-08-400-256-47

Alignment Scores:
Pred. No.: 3 09 Length: 538
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-400-256-47 (1-538)

QY 54 SerAlaSerAlaSerAlaSerLeu 62
Db 354 TCAGCTTCAGCTTCAGCTTCCTC 328

RESULT 7
US-08-975-365-47/c
Sequence 47, Application US/08975365
Patent No. 6011007
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Asger Sloth
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985.220-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 77..514
US-08-975-365-47

Alignment Scores:
Pred. No.: 3 09 Length: 538
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 3 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-975-365-47 (1-538)

QY 54 SerAlaSerAlaSerAlaSerLeu 62
Db 354 TCAGCTTCAGCTTCAGCTTCCTC 328

RESULT 8
US-08-400-256-12/c
Sequence 12, Application US/08400256
Patent No. 5750497
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Asger Sloth
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 03-MAR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985.220-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-400-256-12

Alignment Scores:
Pred. No.: 3.91 Length: 63
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.69% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-400-256-12 (1-63)

Qy 55 AlasSerAlaSerLeu 62
Db 44 GCCTCAGCTTCAGCTTCTCTC 21

RESULT 9

US-08-975-365-12/c
Sequence 12, Application US/08975365
Patent No. 6011007
GENERAL INFORMATION:
APPLICANT: Havelund, Svend
APPLICANT: Halstrom, John
APPLICANT: Jonassen, Ib
APPLICANT: Andersen, Asger Sloth
APPLICANT: Markussen, Jan
TITLE OF INVENTION: ACYLATED INSULIN
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,365
FILING DATE:

CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/400,256
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3985.220-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 63 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

US-08-975-365-12

Alignment Scores:
Pred. No.: 3.91 Length: 63
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.69% Indels: 0
DB: 3 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-975-365-12 (1-63)

Qy 55 AlasSerAlaSerLeu 62
Db 44 GCCTCAGCTTCAGCTTCTCTC 21

RESULT 10

US-09-097-767A-8
Sequence 8, Application US/09097767A
Patent No. 6169232
GENERAL INFORMATION:
APPLICANT: Hey, Timothy D.
APPLICANT: Owens Merio, Ann
APPLICANT: Walsh, Terence A.
TITLE OF INVENTION: Nucleotide Sequences of Genes
TITLE OF INVENTION: Encoding Sink Proteins and Uses Thereof for Improving the N
TITLE OF INVENTION: Quality of Feeds
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences, LLC
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0,
SOFTWARE: Version 1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,767A
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Borucki, Andrea T.
REGISTRATION NUMBER: 33651
REFERENCE/DOCKET NUMBER: 50451
TELEPHONE: 317-337-4846
TELEFAX: 317-337-4847

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 72 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-097-767A-8

Alignment Scores:

Pred. No.: 4.41 Length: 72
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.69% Indels: 0
DB: 4 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-097-767A-8 (1-72)

Qy 54 SerAlaSerAlaSerLeu 61
Db 44 GCCTCAGCTTCAGCTTCTCTC 21

Db 23 TCGCATCTGCATCTGCATCTTCA 46

RESULT 11

US-09-097-767A-7/C

; Sequence 7, Application US/09097767A

; Patent No. 6169232

; GENERAL INFORMATION:

; APPLICANT: Hey, Timothy D.

; APPLICANT: Owens Merlo, Ann

; APPLICANT: Walsh, Terence A.

; TITLE OF INVENTION: Nucleotide Sequences of Genes

; TITLE OF INVENTION: Encoding Sink Proteins and Uses Thereof for Improving the Nutri

; TITLE OF INVENTION: Quality of Feeds

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dow Agrosciences, LLC

; STREET: 9330 Zionsville Road

; CITY: Indianapolis

; STATE: IN

; COUNTRY: US

; ZIP: 46268

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0,

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/097,767A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Borucki, Andrea T.

; REGISTRATION NUMBER: 33651

; REFERENCE/DOCKET NUMBER: 50451

; TELEPHONE: 317-337-4846

; TELEFAX: 317-337-4847

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 79 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

US-09-097-767A-7

Alignment Scores: 4.8 Length: 79

Pred. No.: 8

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 7.69% Indels: 0

DB: 4 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-097-767A-7 (1-79)

QY 54 SerAlaSerAlaSerAlaSerSer 61

Db 53 TCGGCATCTGCATCTGCATCTTCA 30

RESULT 12

US-08-468-674B-64

; Sequence 64, Application US/08468674B

; Patent No. 5639642

; GENERAL INFORMATION:

; APPLICANT: Kjeldsen, Thomas B

; APPLICANT: Vad, Knud

; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES

; NUMBER OF SEQUENCES: 89

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5639642o No. 5639642th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Tape

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/780,571

; FILING DATE: 08-JAN-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/468,674

; FILING DATE: 06-JUN-1995

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Tape

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/468,674B

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/282,852

; FILING DATE: 29-JUL-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4085.220-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 64:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 82 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-468-674B-64

Alignment Scores: 4.96 Length: 82

Pred. No.: 8.00 Matches: 8

Score: 8.00 Conservative: 0

Percent Similarity: 100.00% Mismatches: 0

Best Local Similarity: 100.00% Indels: 0

Query Match: 7.69% Gaps: 0

DB: 1

US-09-814-661A-2 (1-104) x US-08-468-674B-64 (1-82)

QY 55 AlaserAlaserAlaserSerLeu 62

Db 18 GCTTCAGCTTCAGCTTCTCTC 41

RESULT 13

US-08-780-571-64

; Sequence 64, Application US/08780571

; Patent No. 5795746

; GENERAL INFORMATION:

; APPLICANT: Kjeldsen, Thomas B

; APPLICANT: Vad, Knud

; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES

; NUMBER OF SEQUENCES: 89

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5795746o No. 5795746th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Tape

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/780,571

; FILING DATE: 08-JAN-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/468,674

; FILING DATE: 06-JUN-1995

; APPLICATION NUMBER: US 08/282,852
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4085.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 82 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-780-571-64

Alignment Scores:
Pred. No.: 4.96 Length: 82
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.69% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-780-571-64 (1-82)
QY 55 AlaserAlaserSerLeu 62
Db 18 GCTTCAGCTTCAGCTTCTCTC 41

RESULT 14
US-08-883-070-2
; Sequence 2, Application US/08883070
; Patent No. 598599
; GENERAL INFORMATION:
; APPLICANT: Ruegg, C.L. et al.
; TITLE OF INVENTION: GROWTH ARREST GENE COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306-1546
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,070
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/021,874
; FILING DATE: 28-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 7636-0011.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: B4B
; US-08-883-070-2

Alignment Scores:
Pred. No.: 23.9 Length: 471
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.69% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-2 (1-104) x US-08-883-070-2 (1-471)
QY 54 SerAlaserAlaserSer 61
Db 416 TCTGCTTCGCTTCAGCTTCATCA 439

RESULT 15
US-09-097-767A-36/c
; Sequence 36, Application US/09097767A
; Patent No. 6169232
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy D.
; APPLICANT: Owens Merlo, Ann
; APPLICANT: Walsh, Terence A.
; TITLE OF INVENTION: Nucleotide Sequences of Genes
; TITLE OF INVENTION: Encoding Sink Proteins and Uses Thereof for Improving the N
; TITLE OF INVENTION: Quality of Feeds
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences, LLC
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,767A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Borucki, Andrea T.
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 50451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-337-4846
; TELEFAX: 317-337-4847
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-097-767A-36

Alignment Scores:
Pred. No.: 26.7 Length: 533
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.69% Indels: 0
DB: 4 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-097-767A-36 (1-533)

OY 54 SerAlaSerAlaSerAlaSerSer 61
|||||
Db 148 TCGGCATCTGCATCTGCATCTCA 125

Search completed: July 17, 2003, 20:23:05
Job time : 67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 19:00:20 ; Search time 213 Seconds
(without alignments)
1099.567 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 104

Sequence: 1 MQNSQDYFYAQRNCRQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368727

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09814661/runat_16072003_115400_15615/app_query.fasta_1.263
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661.ecgn.1.1.396.ernat_16072003_115400_15615 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	104	100.0	1158	21	AAA14263	Yeast SML1 (suppre
c 2	10	9.6	3423	23	ABL24497	Drosophila melanog
c 3	10	9.6	6222	23	ABL24496	Drosophila melanog
c 4	9	8.7	65	21	AAA14269	PCR primer SML1-HI
c 5	9	8.7	72	17	AAT30945	PI-primer for N-te
c 6	9	8.7	74	17	AAT30946	PI-primer #4 for N
c 7	9	8.7	83	17	AAT30947	PI-primer #5 for N
c 8	9	8.7	112	17	AAT39029	PCR primer 593 for
c 9	9	8.7	219	17	AAT10544	N-terminally exten
c 10	9	8.7	360	21	AAC56572	Eucalyptus grandis
c 11	9	8.7	379	17	AAT10547	PAK625 DNA fragmen
c 12	9	8.7	450	24	ABK65281	Arabidopsis cDNA e
c 13	9	8.7	538	16	AAO86432	Mating factor alph
c 14	9	8.7	1275	23	AAS81350	DNA encoding novel
c 15	9	8.7	1556	23	ABL06789	Drosophila melanog
c 16	9	8.7	1603	22	AAH21932	Mouse total gene e
c 17	9	8.7	1603	24	ABK71489	Murine CLZ-58 - OG
c 18	9	8.7	4054	23	ABL15295	Drosophila melanog
c 19	9	8.7	4275	23	ABL06788	Drosophila melanog
c 20	9	8.7	4541	23	ABL24072	Drosophila melanog
c 21	9	8.7	6238	23	ABL15294	Drosophila melanog
c 22	9	8.7	6917	23	ABL17935	Drosophila melanog
c 23	9	8.7	10031	23	ABL17934	Drosophila melanog
c 24	8	7.7	60	24	ABN46188	Human spliced tran
c 25	8	7.7	63	16	AAO86420	Human insulin deri
c 26	8	7.7	63	17	AAT30942	PI-primer #1 for N
c 27	8	7.7	64	17	AAT30943	PI-primer #2 for N
c 28	8	7.7	72	20	AAK78469	Maize RIP-2 PCR pr
c 29	8	7.7	79	20	AAK78468	Maize RIP-2 PCR pr
c 30	8	7.7	82	17	AAT10525	MI3 insulin precu
c 31	8	7.7	84	21	AAC31994	Human secreted pro
c 32	8	7.7	153	23	AAI84452	Human polynucleoti
c 33	8	7.7	243	6	AAO50112	Sequence encoding
c 34	8	7.7	243	8	AAH71117	Novel receptor pro
c 35	8	7.7	369	21	AAZ92677	DNA encoding Haemo
c 36	8	7.7	372	23	AAS53338	Haemophilus influe
c 37	8	7.7	381	21	AAC07458	Human secreted pro
c 38	8	7.7	429	21	AAC02027	Human secreted pro
c 39	8	7.7	431	24	ABK76196	Bacillus lichenifo
c 40	8	7.7	435	23	ABL14807	Drosophila melanog
c 41	8	7.7	456	24	ABQ56996	Human colon cancer
c 42	8	7.7	465	21	AAC01520	Human secreted pro
c 43	8	7.7	466	21	AAC42435	Arabidopsis thalia
c 44	8	7.7	471	23	ABL27151	Drosophila melanog
c 45	8	7.7	527	21	AAC04030	Human secreted pro

ALIGNMENTS

RESULT 1

AAA14263

ID AAA14263 standard; DNA; 1158 BP.

XX AAA14263;

AC AAA14263;

XX 15-AUG-2000 (first entry)

DE Yeast SML1 (suppressor of mecl lethality) gene.

XX SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;

KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;

KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;

KW cell cycle checkpoint; cancer; microbial infection;

XX ataxia telangiectasia; AT; ds.

XX Saccharomyces cerevisiae.

OS

XX

FH key Location/Qualifiers
 FT CDS 549..863
 FT /*tag= a
 FT /product= "yeast Sml1 protein"
 FT /function= "Inhibitor of ribonucleotide reductase"
 XX
 PN WO200017225-A2.
 XX
 XX
 PD 30-MAR-2000.
 XX
 PF 24-SEP-1999; 99WO-US22260.
 XX
 PR 24-SEP-1998; 98US-0158858.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 XX Rothstein R, Zhao X;
 PI
 DR WPI: 2000-283539/24.
 DR P-PSDB: AAY90470.
 XX
 XX New Sml1 protein and its homologs, useful for treating cancer,
 PT microbial infection and ataxia telangiectasia and in screening for
 PT specific modulators
 XX
 PS Claim 6; Fig 1C; 98pp; English.
 XX
 CC This sequence represents the yeast SML1 (suppressor of mecl lethality)
 CC gene. This gene, located on chromosome XIII, encodes Sml1, an inhibitor
 CC of ribonucleotide reductase. The SML1 gene (assigned the name YML058W by
 CC the Stanford yeast genomic project) was initially isolated in a mecl-1
 CC strain by genetic analysis. Sml1 is a suppressor of the mecl mutant in
 CC yeast which is associated with abnormal levels of recombination in both
 CC meiosis and mitosis. Sml1 also permits cell growth in the absence of the
 CC cell cycle checkpoint proteins Mec1 and Rad53, which, unlike most
 CC checkpoint proteins, are essential for cell growth. Sml1 binds to
 CC the largest subunit of ribonucleotide reductase (RNR) which catalyses the
 CC rate-limiting step of deoxynucleotide triphosphate (dNTP) synthesis,
 CC conversion of NTPs to dNTPs. Binding of Sml1 reduces the activity of RNR
 CC and thus inhibits the synthesis of dNTPs. The SML1 gene, or fragments
 CC thereof may be used to isolate the human homologue of SML1. The Sml1
 CC protein, and compounds that modulate the interaction of Sml1 with
 CC ribonucleotide reductase (RNR), may be used to alter the rate at which
 CC cells divide. These are particularly useful for treating cancer,
 CC microbial infection and ataxia telangiectasia (AT), a condition in which
 CC patients are prone to cancer. Sml1 may also be used to screen for
 CC modulatory agents, to raise specific antibodies, and for stimulating the
 CC function of the ATM gene (a mammalian Mec1 homologue which is mutated in
 CC AT). Anti-Sml1 antibodies are used as diagnostic and analytical
 CC immunoassay reagents and to remove Sml1 from serum or to titrate Sml1
 CC intracellularly.
 XX
 SQ Sequence 1158 BP; 289 A; 282 C; 196 G; 391 T; 0 other;

Alignment Scores:
 Pred. No.: 1.68e-92 Length: 1158
 Score: 104.00 Matches: 104
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-814-661A-2 (1-104) x AAA14263 (1-1158)
 OY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 DB 549 ATGCAGAAATATCCCAAGACTACTTTTACGCTCAAAATCGCTGCCAACACACAGCCCT 609
 OY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProProMetAla 40
 DB 609 TCCACATTCGCTACCGTGACCATCGCGGATTTAGAGGGTGCTTGGCCACCTATGGCT 668
 OY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerAlaSerAlaSer 60

Db 669 GAGGTTCTTATGTGTCTACTCAAACTCCATGGCAGCTCCGCTTCTCGCTTCT 728
 OY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
 Db 729 TCATTAGAAATGTGGGAAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 788
 OY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
 Db 789 AACAAACAACAATTTGGTCTTGGCGAACTAAATCTATGTTCACACGAGGTAGGTGCG 848
 OY 101 GluMetAspPhe 104
 Db 849 GAAATGACCTTC 860
 RESULT 2
 ABL24497/C
 ID ABL24497 standard; DNA: 3423 BP.
 XX
 AC ABL24497;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24964.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 24964; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3423 BP; 955 A; 958 C; 905 G; 605 T; 0 other;

Alignment Scores:
 Pred. No.: 5.59 Length: 3423
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.62% Indels: 0
 DB: 23 Gaps: 0

US-09-814-661A-2 (1-104) x ABL24497 (1-3423)

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
 ID |||||||
 Db 2057 TCCTCCGCTCGGCTCGGCATCTCCTTA 2028

RESULT 3

ABL24496
 ID ABL24496 standard; DNA; 6222 BP.

XX AC ABL24496;

DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24961.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PERE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

PS Claim 1; SEQ ID NO 24961; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 6222 BP; 1519 A; 1418 C; 1452 G; 1833 T; 0 other;

Alignment Scores:

Pred. No.:	9,88	Length:	6222
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	9.62%	Indels:	0
DB:	23	Gaps:	0

US-09-814-661A-2 (1-104) x ABL24496 (1-6222)

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62

Db 2449 TCCTCCGCTCGGCTCGGCATCTCCTTA 2478

RESULT 4

AA14269/C

ID AA14269 standard; DNA; 65 BP.

XX

AC AA14269;

XX 15-AUG-2000 (first entry)

XX DE PCR primer SML1-HIS3 3', used in the study of the yeast SML1 gene.

XX KW SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;
 KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
 KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
 KW cell cycle checkpoint; cancer; microbial infection;
 KW ataxia telangiectasia; AT; PCR primer; ss.

XX OS Saccharomyces cerevisiae.

XX PN WO200017225-A2.

XX PD 30-MAR-2000.

XX PF 24-SEP-1999; 99WO-US22260.

XX PR 24-SEP-1998; 98US-0158858.

XX PA (UYCO) UNIV COLUMBIA NEW YORK.

XX PI Rothstein R, Zhao X;

XX DR WPI; 2000-283539/24.

XX PT New Sml1 protein and its homologs, useful for treating cancer,
 PT microbial infection and ataxia telangiectasia and in screening for
 PT specific modulators.

XX PS Disclosure; Page 65; 98pp; English.

XX CC The invention relates to the yeast SML1 (suppressor of mecl lethality)
 CC gene (AA14263) and to the Sml1 protein (AA90470). The Sml1 gene,
 CC located on chromosome XIII, encodes Sml1, an inhibitor of ribonucleotide
 CC reductase. The SML1 gene (assigned the name YML058w by the Stanford yeast
 CC genomic project) was initially isolated in a mecl-1 strain by genetic
 CC analysis. Sml1 is a suppressor of the mecl mutant in yeast which is
 CC associated with abnormal levels of recombination in both meiosis and
 CC mitosis. Sml1 also permits cell growth in the absence of the cell cycle
 CC checkpoint proteins Mec1 and Rad53, which, unlike most checkpoint
 CC proteins, are essential for cell growth. Sml1 binds to the largest
 CC subunit of ribonucleotide reductase (RNR) which catalyses the
 CC rate-limiting step of deoxynucleotide triphosphate (dNTP) synthesis,
 CC conversion of NTPs to dNTPs. Binding of Sml1 reduces the activity of RNR
 CC and thus inhibits the synthesis of dNTPs. The SML1 gene, or fragments
 CC thereof may be used to isolate the human homologue of SML1. The Sml1
 CC protein, and compounds that modulate the interaction of Sml1 with
 CC ribonucleotide reductase (RNR), may be used to alter the rate at which
 CC cells divide. These are particularly useful for treating cancer,
 CC microbial infection and ataxia telangiectasia (AT), a condition in which
 CC patients are prone to cancer. Sml1 may also be used to screen for
 CC modulatory agents, to raise specific antibodies, and for stimulating the
 CC function of the ATM gene (a mammalian Mec1 homologue which is mutated in
 CC AT). Anti-Sml1 antibodies are used as diagnostic and analytical
 CC immunoassay reagents and to remove Sml1 from serum or to titrate Sml1
 CC intracellularly. Sequences AA24266-AA14285 represent PCR primers used in
 CC the study of the yeast SML1 gene.

XX SQ Sequence 65 BP; 17 A; 17 C; 14 G; 17 T; 0 other;

Alignment Scores:

Pred. No.:	1,22	Length:	65
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.65%	Indels:	0
DB:	21	Gaps:	0

US-09-814-661A-2 (1-104) x AA14269 (1-65)

QY 96 GlnGlyLysValGluGluMetAspPhe 104
 |||||
 DB 44 CAGGTAAGTCGAGGAATGCACTTC 18

RESULT 5
 AAT30945/c
 ID AAT30945 standard; DNA; 72 BP.
 XX
 AC AAT30945;
 XX
 DT 08-AUG-1996 (first entry)
 XX
 DE Pl-primer for N-terminally extended glucagon-like peptide-1.
 XX
 KW Signal peptide; alpha-factor; dipeptidyl aminopeptidase processing; PCR;
 KW proteolysis; yeast; aprotinin; tissue factor pathway inhibitor; ILGF-I;
 KW protease inhibitor; insulin-like growth factor I; ILGF-II; interleukin;
 KW human growth hormone; bovine growth hormone; glucagon-like peptide-1;
 KW tissue plasminogen activator; glucagon; Factor VII; Factor VIII; insulin;
 KW Factor XIII; platelet derived growth factor; insulin precursor; primer;
 KW amplify; polymerase chain reaction; ss.
 XX
 OS Synthetic.
 XX
 PN W09535384-A1.
 XX
 PD 28-DEC-1995.
 XX
 PF 16-JUN-1995; 95WO-DK00250.
 XX
 PR 17-JUN-1994; 94DK-0000712.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Brandt J, Kjeldsen TB, Vad K;
 DR WPI; 1996-058421/06.
 XX
 PT DNA construct for expressing N-terminally extended heterologous
 PT protein in yeast - enabling higher yield of correctly processed
 PT proteins such as growth hormone, interleukin, tissue plasminogen
 PT activator and insulin
 XX
 PS Example 4; Page 22; 79pp; English.
 XX
 CC This sequence represents the PI-primer sequence, used in conjunction
 CC with the sequence represented by AAT30941 to amplify the N-terminally
 CC extended glucagon-like peptide-1. The amplified sequence is used as an
 CC N-terminally extended heterologous protein (HP) coding sequence. The
 CC coding sequences for a signal peptide, a leader sequence and a yeast
 CC processing site are attached to this sequence. This construct is then
 CC inserted into a recombinant expression vector which is used to transform
 CC a yeast strain capable of expressing the heterologous protein. The
 CC yeast processing site, contained within the N-terminal extension, enables
 CC the leader sequence and signal peptide to be removed from the expressed
 CC protein. The N-terminal extension of the HP increases the fermentation
 CC yield and protects against dipeptidyl aminopeptidase processing. The
 CC extension can be cleaved off either by naturally occurring yeast
 CC proteases before purification from the culture media or by in vitro
 CC proteolysis during or subsequent to purification of the product. These
 CC constructs can be used to obtain higher yields of correctly processed
 CC proteins such as aprotinin, tissue factor pathway inhibitor, protease
 CC inhibitors, insulin-like growth factor I (ILGF-I), ILGF-II, human or
 CC bovine growth hormone, interleukin, tissue plasminogen activator,
 CC glucagon, glucagon-like peptide-1, Factor VII, Factor VIII, Factor XIII,
 CC platelet derived growth factor, enzymes, insulin, or an insulin
 CC precursor.
 XX
 SQ Sequence 72 BP; 24 A; 15 C; 21 G; 12 T; 0 other;

Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AAT30945 (1-72)
 QY 54 SerAlaSerAlaSerAlaSerSerleu 62
 |||||
 DB 56 TCAGCTTCAGCTTCAGCTTCCTCTC 30

RESULT 6
 AAT30946/c
 ID AAT30946 standard; DNA; 74 BP.
 XX
 AC AAT30946;
 XX
 DT 08-AUG-1996 (first entry)
 XX
 DE Pl-primer #4 for N-terminally extended insulin precursor.
 XX
 KW Signal peptide; alpha-factor; dipeptidyl aminopeptidase processing; PCR;
 KW proteolysis; yeast; aprotinin; tissue factor pathway inhibitor; ILGF-I;
 KW protease inhibitor; insulin-like growth factor I; ILGF-II; interleukin;
 KW human growth hormone; bovine growth hormone; glucagon-like peptide-1;
 KW tissue plasminogen activator; glucagon; Factor VII; Factor VIII; insulin;
 KW Factor XIII; platelet derived growth factor; insulin precursor; primer;
 KW amplify; polymerase chain reaction; ss.
 XX
 OS Synthetic.
 XX
 PN W09535384-A1.
 XX
 PD 28-DEC-1995.
 XX
 PF 16-JUN-1995; 95WO-DK00250.
 XX
 PR 17-JUN-1994; 94DK-0000712.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Brandt J, Kjeldsen TB, Vad K;
 DR WPI; 1996-058421/06.
 XX
 PT DNA construct for expressing N-terminally extended heterologous
 PT protein in yeast - enabling higher yield of correctly processed
 PT proteins such as growth hormone, interleukin, tissue plasminogen
 PT activator and insulin
 XX
 PS Example 9; Page 27; 79pp; English.
 XX
 CC This sequence represents the PI-primer sequence, used in conjunction
 CC with the sequence represented by AAT30941 to amplify the N-terminally
 CC extended insulin precursor protein coding sequence. The amplified
 CC sequence is used as an N-terminally extended heterologous protein (HP)
 CC coding sequence. The coding sequences for a signal peptide, a leader
 CC sequence and a yeast processing site are attached to this sequence. This
 CC construct is then inserted into a recombinant expression vector which is
 CC used to transform a yeast strain capable of expressing the heterologous
 CC protein. The yeast processing site, contained within the N-terminal
 CC extension, enables the leader sequence and signal peptide to be removed
 CC from the expressed protein. The N-terminal extension of the HP increases
 CC the fermentation yield and protects against dipeptidyl aminopeptidase
 CC processing. The extension can be cleaved off either by naturally
 CC occurring yeast proteases before purification from the culture media or
 CC by in vitro proteolysis during or subsequent to purification of the
 CC product. These constructs can be used to obtain higher yields of
 CC correctly processed proteins such as aprotinin, tissue factor pathway
 CC inhibitor, protease inhibitors, insulin-like growth factor I (ILGF-I),
 CC ILGF-II, human or bovine growth hormone, interleukin, tissue plasminogen
 CC activator, glucagon, glucagon-like peptide-1, Factor VII, Factor VIII,

Alignment Scores:
 Pred. No.: 1.34 Length: 72

CC Factor XIII, platelet derived growth factor, enzymes, insulin, or an
 CC insulin precursor.

SQ Sequence 74 BP; 19 A; 19 C; 22 G; 14 T; 0 other;

Alignment Scores:

Pred. No.: 1.38 Length: 74
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AAT30946 (1-74)

QY 54 SerAlaSerAlaSerAlaSerSerLeu 62

DB 44 TCAGCGTCAGCGTCAGCGTCGTCCTC 18

RESULT 7

AAT30947/C

ID AAT30947 standard; DNA; 83 BP.

XX AC AAT30947;

XX DT 08-AUG-1996 (first entry)

XX DE Pl-primer #5 for N-terminally extended insulin precursor.

XX KW Signal peptide; alpha-factor; dipeptidyl aminopeptidase processing; PCR;
 KW proteolysis; yeast; aprotinin; tissue factor pathway inhibitor; ILGF-I;
 KW protease inhibitor; insulin-like growth factor I; ILGF-II; interleukin;
 KW human growth hormone; bovine growth hormone; glucagon-like peptide-1;
 KW tissue plasminogen activator; glucagon; Factor VII; Factor VIII; insulin;
 KW Factor XIII; platelet derived growth factor; insulin precursor; primer;
 KW amplify; polymerase chain reaction; ss.

XX OS Synthetic.

XX PN W09535384-A1.

XX PD 28-DEC-1995.

XX PF 16-JUN-1995; 95WO-DK00250.

XX PR 17-JUN-1994; 94DK-0000712.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Brandt J, Kjeldsen TB, Vad K;

XX DR WPI; 1996-058421/06.

XX PT DNA construct for expressing N-terminally extended heterologous
 PT protein in yeast - enabling higher yield of correctly processed
 PT proteins such as growth hormone, interleukin, tissue plasminogen
 PT activator and insulin

XX PS Example 10; Page 27; 79pp; English.

XX CC This sequence represents the Pl-primer sequence, used in conjunction
 CC with the sequence represented by AAT30941 to amplify the N-terminally
 CC extended insulin precursor protein coding sequence. The amplified
 CC sequence is used as an N-terminally extended heterologous protein (HP)
 CC coding sequence. The coding sequences for a signal peptide, a leader
 CC sequence and a yeast processing site are attached to this sequence. This
 CC construct is then inserted into a recombinant expression vector which is
 CC used to transform a yeast strain capable of expressing the heterologous
 CC protein. The yeast processing site, contained within the N-terminal
 CC extension, enables the leader sequence and signal peptide to be removed
 CC from the expressed protein. The N-terminal extension of the HP increases
 CC the fermentation yield and protects against dipeptidyl aminopeptidase
 CC processing. The extension can be cleaved off either by naturally

CC occurring yeast proteases before purification from the culture media or
 CC by in vitro proteolysis during or subsequent to purification of the
 CC product. These constructs can be used to obtain higher yields of
 CC correctly processed proteins such as aprotinin, tissue factor pathway
 CC inhibitor, protease inhibitors, insulin-like growth factor I (ILGF-I),
 CC ILGF-II, human or bovine growth hormone, interleukin, tissue plasminogen
 CC activator, glucagon, glucagon-like peptide-1, Factor VII, Factor VIII,
 CC Factor XIII, platelet derived growth factor, enzymes, insulin, or an
 CC insulin precursor.

SQ Sequence 83 BP; 28 A; 16 C; 24 G; 15 T; 0 other;

Alignment Scores:

Pred. No.: 1.54 Length: 83
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AAT30947 (1-83)

QY 54 SerAlaSerAlaSerAlaSerSerLeu 62

DB 44 TCAGCTTCAGCTTCAGCTTCCTCTC 18

RESULT 8

AAT30929

ID AAT30929 standard; DNA; 112 BP.

XX AC AAT30929;

XX DT 12-MAY-1997 (first entry)

XX DE PCR primer 593 for insulin precursor encoding plasmid pMT742.

XX KW pMT742; prepro-leader; precursor; Saccharomyces cerevisiae;
 KW yeast strain; YKV153; lipophilic group; insulin; derivative;
 KW treatment; diabetes; soluble; physiological pH; primer; PCR;
 KW protracted action profile; polymerase chain reaction; ss.

XX OS Synthetic.

XX PN W09629344-A1.

XX PD 26-SEP-1996.

XX PF 18-MAR-1996; 96WO-DK00107.

XX PR 17-MAR-1995; 95DK-0000276.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Brandt J, Hansen PH, Havelund S, Jonassen I, Kaarsholm NC;
 PI Kurtzhals P, Markussen J;

XX DR WPI; 1996-443135/44.

XX PT New insulin derivs. with lipophilic gps. - having a protracted
 PT profile of action at physiological pH, used for treating diabetes.

XX PS Example 1; Page 22; 58pp; English.

XX CC The present sequence is a PCR primer for the plasmid pMT742,
 CC which was used for the expression of an insulin precursor from
 CC yeast strain YKV153 using the Saccharomyces cerevisiae MF alpha
 CC prepro-leader. The insulin precursor was used in the preparation of
 CC claimed lipophilic group containing insulin derivatives, which can
 CC be used for the treatment of diabetes, are soluble at physiological
 CC pH values and have a protracted action profile.

XX SQ Sequence 112 BP; 31 A; 29 C; 20 G; 32 T; 0 other;

Alignment Scores:
 Pred. No.: 2.05 Length: 112
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AAT39029 (1-112)

QY 54 SerAlaSerAlaSerAlaSerLeu 62
 ID AAT10544/c
 DB 69 TCAGCTTCAGCTTCAGCTTCCTCT 95

RESULT 9

AAT10544/c
 ID AAT10544 standard; DNA; 219 BP.

XX AC AAT10544;

DT 25-JUL-1996 (first entry)

DE N-terminally extended M13 insulin precursor DNA.

KW Leader sequence; M13; insulin precursor;

KW expression; secretion; cassette; alpha-factor;

KW mouse salivary amylase; carboxypeptidase; yeast;

KW aspartic protease 3; BAR1; KEX2 protease; ds.

OS Synthetic.

XX XX

FH Key Location/Qualifiers

FT Mat_peptide 1..195
 /*tag= a

XX W09534666-A1.

PD 21-DEC-1995.

XX PF 16-JUN-1995; 95WO-DK00249.

XX PR 29-JUL-1994; 94US-0282852.

XX PR 16-JUN-1994; 94DK-0000705.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Kjeldsen TB, Vad K;

XX DR WPI: 1996-049693/05.

XX DR P-PSDB: AAR88188.

XX Expression cassette for yeast contg. synthetic leader sequence -
 providing high yields of secreted polypeptide encoded by the
 cassette, also related vectors and transformed yeast cells

XX Example 10; Fig 19; 85pp; English.

XX The present sequence encodes the N-terminally extended M13 insulin
 precursor (IP). An expression cassette for the M13 IP in yeast
 comprises 5'-3', a promoter (P), sequences encoding a signal peptide
 (SP), a leader sequence, a processing site (PS) and the M13 IP and
 an optional terminator sequence. The P can be any P functional in
 yeast, e.g. the alpha-factor gene P, and the SP is pref. the alpha
 factor, mouse salivary amylase, carboxypeptidase, yeast aspartic
 protease 3 or yeast BAR1 SP. The PS is LysArg, ArgLys, ArgArg or
 LysLys, for processing by *S. cerevisiae* KEX2 protease.

XX Sequence 219 BP; 59 A; 49 C; 52 G; 59 T; 0 other;

Alignment Scores:
 Pred. No.: 3.88 Length: 219
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AAT10544 (1-219)

QY 54 SerAlaSerAlaSerAlaSerLeu 62

DB 29 TCAGCTTCAGCTTCAGCTTCCTCTCTC 3

RESULT 10

AAC56572/c

ID AAC56572 standard; DNA; 360 BP.

XX AC AAC56572;

DT 25-JAN-2001 (first entry)

DE Eucalyptus grandis transcription factor DNA sequence #443.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;

KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;

KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;

KW type 2 Cys2His2; CCAAT box element; MYB; ss.

XX Eucalyptus grandis.

OS WO200053724-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06112.

XX PR 11-MAR-1999; 99US-0266513.

XX PR 18-AUG-1999; 99US-0149485.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Wood M, McGrath A, Shenk MA, Glenn M;

XX DR WPI: 2000-579369/54.

XX New isolated polynucleotide encoding a plant transcription factor for
 producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 having modified gene expression or modified activity of a polypeptide

XX Claim 1; Page 466; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 sequence for one such transcription factor. The transcription factor may
 be used to produce a plant having modified gene expression such as a
 woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 mahogany species or to modify the activity of a polypeptide in a plant.
 The transcription factors of the present invention are members from the
 following families of regulatory proteins: bZIP, bZIP family of G-box
 binding factors, basic helix-loop-helix zipper,
 homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 and MYB.

XX Sequence 360 BP; 86 A; 87 C; 126 G; 61 T; 0 other;

Alignment Scores:
 Pred. No.: 6.23 Length: 360
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 21 Gaps: 0

US-09-814-661A-2 (1-104) x AAC56572 (1-360)

QY 53 SerSerAlaSerAlaSerAlaSerSer 61
 Db 79 TCATCGGCATCGGCATCGGCATCGICA 53

RESULT 11
 AAT10547/c
 ID AAT10547 standard; DNA; 379 BP.
 XX
 AC AAT10547;
 XX
 XX 25-JUL-1996 (first entry)
 XX
 XX pAK625 DNA fragment leader sequence construction template.
 DE
 XX Leader sequence; M13; insulin precursor; Saccharomyces cerevisiae;
 KW strain; yAK680; expression; secretion; cassette; alpha-factor;
 KW mouse salivary amylase; carboxypeptidase; yeast; plasmid pAK625;
 KW aspartic protease 3; BARI; KEX2 protease; direct template; ds.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH 113..376
 CDS /*tag= a
 FT /*note= "5'-end fragment only"
 FT
 FT
 FT
 XX W09534666-A1.
 PN
 XX 21-DEC-1995.
 PD
 XX 16-JUN-1995; 95WO-DK00249.
 PF
 XX 29-JUL-1994; 94US-0282852.
 PR
 PR 16-JUN-1994; 94DK-0000705.
 PR
 XX (NOVO) NOVO-NORDISK AS.
 PA
 XX Kjeldsen TB, Vad K;
 PI
 XX WPI; 1996-049693/05.
 DR P-PSDB; AAR88190.
 DR
 XX Expression cassette for yeast contg, synthetic leader sequence
 PT providing high yields of secreted polypeptide encoded by the
 PT cassette, also related vectors and transformed yeast cells
 PT
 XX
 PS Example 11; Fig 22; 85pp; English.
 XX
 CC The present sequence is a DNA fragment of plasmid pAK625, which
 CC was used as the direct template in the construction of the M13
 CC insulin precursor (IP) leader sequence (US) AAR88178. The LS
 CC is used to express the M13 IP in S. cerevisiae strain YAK680,
 CC providing high level expression and secretion. An expression
 CC cassette for the M13 IP in yeast, comprises 5'-3' a promoter
 CC (P), sequences encoding a signal peptide (SP), a leader
 CC sequence, a processing site (PS) and the M13 IP and an optional
 CC terminator sequence. The P can be any P functional in yeast, e.g.
 CC the alpha-factor gene P, and the SP is pref. the alpha-factor,
 CC mouse salivary amylase, carboxypeptidase, yeast aspartic
 CC protease 3 or yeast BARI SP. The PS is LysArg, ArgLys, ArgArg or
 CC LysLys, for processing by S. cerevisiae KEX2 protease.
 XX
 SQ Sequence 379 BP; 129 A; 81 C; 63 G; 106 T; 0 other;

Alignment Scores:
 Pred. No.: 6.54 Length: 379
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0

DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AAT10547 (1-379)

QY 54 SerAlaSerAlaSerAlaSerSerLeu 62
 Db 321 TCAGCTTCAGCTTCAGCTTCCTCTCTC 295

RESULT 12
 ABK65281
 ID ABK65281 standard; cDNA; 450 BP.
 XX
 AC ABK65281;
 XX
 XX 02-JUL-2002 (first entry)
 DT
 XX
 DE Arabidopsis cDNA encoding a transcription factor #133.
 XX
 KW Plant; ss; gene; transcription factor; transgenic;
 KW agriculture; metabolic chemical; environmental stress; drought;
 KW microbial disease resistance; herbicide resistance; seed yield;
 KW fruit yield; growth rate; leaf senescence; flower senescence.
 XX
 OS Arabidopsis thaliana.
 XX
 XX W0200215675-A1.
 PN
 XX 28-FEB-2002.
 PD
 XX 22-AUG-2001; 2001WO-US26189.
 PF
 XX 22-AUG-2000; 2000US-227439P.
 PR
 PR 16-NOV-2000; 2000US-0713994.
 PR
 PR 16-APR-2001; 2001US-0837944.
 PR
 XX (MEND-) MENDEL BIOTECHNOLOGY INC.
 PA (PILG/) PILGRIM M.
 PA (CREE/) CREELMAN R.
 PA (DUBE/) DUBELL A J.
 PA (HEAR/) HEARD J.
 PA (JIAN/) JIANG C.
 PA (KEDD/) KEDDIE J.
 PA (ADAM/) ADAM L.
 PA (RATC/) RATCLIFF O.
 PA (REUB/) REUBER J L.
 PA (RIEC/) RIECHMANN J L.
 PA (YUGG/) YU G.
 PA (PINE/) PINEDA O.
 XX
 XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;
 PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
 PI
 XX WPI; 2002-292022/33.
 DR P-PSDB; AAU93095.
 DR
 XX An isolated or recombinant polynucleotide used to produce a transgenic
 PT plant ;
 PT
 XX Claim 4; Page 546-547; 94lpp; English.
 PS
 XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
 CC encoding an Arabidopsis thaliana transcription factor, their variants,
 CC complements, fragments, or related polynucleotide with 31% to 95%
 CC sequence identity, where the plant possesses an altered trait as compared
 CC to a wild-type or reference plant, or the plant exhibits an altered
 CC phenotype as compared to a wild-type or reference plant, or the plant
 CC exhibits ectopic expression or altered expression of one or more genes
 CC associated with a plant trait as compared to a wild plant. Also included
 CC are a transgenic plant comprising the polynucleotides, a computer
 CC readable medium having stored sequence information, and identifying a
 CC homologue sequence from a database comprising a plurality of known plant
 CC sequences comprising inputting sequence information selected from one of
 CC 464 fully defined sequences given in the specification. The isolated or

CC recombinant polynucleotide is used for producing a plant having a
 CC modified trait, the method comprising selecting a polynucleotide that
 CC encodes a polypeptide or an antisense nucleic acid, inserting the
 CC polynucleotide or antisense nucleic acid into an expression vector,
 CC introducing the vector into a plant or a cell of a plant to overexpress
 CC the polypeptide or antisense nucleic acid, thereby producing a modified
 CC plant, and selecting for a modified trait (e.g. increased
 CC production of agriculturally useful proteins or metabolic chemicals,
 CC pest tolerance, environmental stress response (e.g. drought), microbial
 CC disease resistance, herbicide resistance, seed and fruit yield, growth
 CC rate, leaf and flower senescence and many other traits listed in the
 CC specification). The present sequence is one of the 232 polynucleotides
 CC encoding an A. thaliana transcription factor.

SQ Sequence 450 BP; 125 A; 92 C; 125 G; 108 T; 0 other;

Alignment Scores:
 Pred. No.: 7.71 Length: 450
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 24 Gaps: 0

US-09-814-661A-2 (1-104) x ABK65281 (1-450)

QY 53 SerSerAlaSerAlaSerAlaSerSer 61
 DB 297 TCCTCGGCTCGGCTCGGCTCTCTCT 323

RESULT 13
 ID AAQ86432/C
 AC AAQ86432 standard; CDNA: 538 BP.

XX
 XX
 XX 20-NOV-1995 (first entry)
 XX
 XX Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.
 XX
 XX Human insulin precursor ArgB1, ArgB31; diabetes; Zinc ion complex;
 KW mating factor alpha 1; N-terminal EEAEAEAE; ss.

XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 77..517
 FT /*tag= a

XX WO9507931-A.

XX 23-MAR-1995.

XX 16-SEP-1994; 94WO-DK00347.

XX 17-SEP-1993; 93DK-0001044.

XX 02-FEB-1994; 94US-0190829.

XX (NOVO) NOVO-NORDISK AS.

XX Andersen AS, Halstrom JB, Havelund S, Jonassen I;

XX Markussen J;

XX WPI: 1995-131314/17.

XX P-PSDB; AAR71695.

XX Acylated insulin deriv. which may be present as a zinc ion
 PT complex - is used to treat diabetes and is rapid acting.

XX Example 6; Page 85; 100pp; English.

XX AAQ86432 encodes AAR71695 mating factor alpha 1-Insulin precursor
 CC ArgB1, ArgB31 N-terminal EEAEAEAE. The insulin precursor comprises

CC the B and A chains of a claimed human insulin derivative preceded
 CC by the N-terminal amino acids EEAEAEAE. In the final claimed compsn.
 CC they are covalently connected via disulphide bonds between Cys
 CC residues A7/B7 and A20/B19. The derivative, which may be present
 CC as a zinc ion complex, can be used as a fast action treatment for
 CC diabetes.

SQ Sequence 538 BP; 168 A; 107 C; 107 G; 156 T; 0 other;

Alignment Scores:
 Pred. No.: 9.14 Length: 538
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.65% Indels: 0
 DB: 16 Gaps: 0

US-09-814-661A-2 (1-104) x AAQ86432 (1-538)

QY 54 SerAlaSerAlaSerAlaSerSerLeu 62
 DB 354 TCAGCTTCAGCTTCAGCTTCCTCTC 328

RESULT 14
 ID AAQ81350
 AC AAQ81350 standard; CDNA: 1275 BP.

XX
 XX
 XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #17154.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang VT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG17163.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX Claim 1; SEQ ID No 17154; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1275 BP; 424 A; 267 C; 332 G; 252 T; 0 other;

Alignment Scores: 20.8 Length: 1275
Pred. No.: 9.00 Matches: 9
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 23 Gaps: 0

US-09-814-661A-2 (1-104) x AAS81350 (1-1275)

OY 53 SerSerAlaSerAlaSerAlaSerSer 61
DB 1229 TCGTCAGCATCAGCATCAGCATCATCC 1255
 |||||

RESULT 15

ABL06789
ID ABL06789 standard; CDNA; 1556 BP.

XX
AC ABL06789;

XX
DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14849.

XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

XX
OS Drosophila melanogaster.

XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.

XX
PF 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX
PA (PEKE) PE CORP NY.

XX
PI Venter JC, Adams M, Li PWD, Myers EW;

XX
XX WPI; 2001-656860/75.

DR P-PSDB; ABB62686.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX
PS Claim 1; SEQ ID NO 14849; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1556 BP; 350 A; 429 C; 442 G; 335 T; 0 other;
Alignment Scores: 25.2 Length: 1556
Pred. No.: 9.00 Matches: 9
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.65% Indels: 0
DB: 23 Gaps: 0

US-09-814-661A-2 (1-104) x ABL06789 (1-1556)

OY 53 SerSerAlaSerAlaSerAlaSerSer 61

DB 1465 TCCAGTGCAGTGCAGTGCAGTGCAGTAGT 1491
 |||||

Search completed: July 17, 2003, 19:59:41
Job time : 217 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 17, 2003, 19:19:04 ; Search time 1314 Seconds
(without alignments)
2303.418 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 104

Sequence: 1 MONSDYFYAQNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09814661/runat_16072003_115400_15623/app_query.fasta_1.263
-DB=GenEmbl -QFW=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn_1.1.3745_@runat_16072003_115400_15623 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.em.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.in.*
18: em.hum.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em_sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	104	100.0	2998	8	SCCMP2
2	104	100.0	20951	8	SC9958
3	10	9.6	3089	8	ACH251521
4	10	9.6	13571	1	AE004982
5	10	9.6	19502	2	AC020305
6	10	9.6	64095	2	AC109752
7	10	9.6	153328	2	AC122111
8	10	9.6	155376	9	AC027026
9	10	9.6	168761	8	AC079887
10	10	9.6	169179	2	AC111888
11	10	9.6	171689	3	AC093435
12	10	9.6	175743	2	AP003579
13	10	9.6	179838	3	AC091220
14	10	9.6	186357	2	AC114454
15	10	9.6	191894	2	AL671875
16	10	9.6	200823	9	AC009652
17	10	9.6	206402	2	AC074335
18	10	9.6	297486	3	AE003550
19	9	8.7	219	6	AR023853
20	9	8.7	219	6	I46879
21	9	8.7	379	6	AR023855
22	9	8.7	379	6	I46881
23	9	8.7	538	6	AR007435
24	9	8.7	626	8	AF041411
25	9	8.7	688	9	HS3334365
26	9	8.7	799	8	AF146868
27	9	8.7	799	8	AF146869
28	9	8.7	799	8	AF146870
29	9	8.7	799	8	AF146871
30	9	8.7	1178	8	AF058796
31	9	8.7	1603	10	MMU066918
32	9	8.7	2112	9	AK092595
33	9	8.7	2166	9	BC027868
34	9	8.7	2360	9	AF193056
35	9	8.7	2535	5	XLJ60093
36	9	8.7	3763	5	PMU19361
37	9	8.7	4053	3	DMU49793
38	9	8.7	4659	1	MTBFSA
39	9	8.7	5135	1	SERDNABP
40	9	8.7	6895	3	DMU09506
41	9	8.7	7388	2	AC014219
42	9	8.7	10029	1	AE012295
43	9	8.7	10348	1	AE009652
44	9	8.7	13486	1	AE004770
45	9	8.7	13695	9	AB065909

ALIGNMENTS

RESULT 1

SCCMP2	SCCMP2	2998 bp	DNA	linear	PLN 25-JUN-1991
LOCUS	Yeast CMP2 gene for calmodulin-binding protein 2.				
DEFINITION	X54964				
ACCESSION	X54964.1	GI:3559			
VERSION	calmodulin binding protein; CMP2 gene.				
KEYWORDS	Saccharomyces cerevisiae.				
SOURCE	Saccharomyces cerevisiae				
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
REFERENCE	1 (bases 1 to 2998)				
AUTHORS	Liu, Y., Ishii, S., Tokai, M., Tsutsumi, H., Ohki, O., Akada, R., Tanaka, K., Tsuchiya, E., Fukui, S. and Miyakawa, T.				
TITLE	The Saccharomyces cerevisiae genes (CMP1 and CMP2) encoding calmodulin-binding proteins homologous to the catalytic subunit of mammalian protein phosphatase 2B				
JOURNAL	Mol. Gen. Genet. 227 (1), 52-59 (1991)				
MEDLINE	91260679				
PUBMED	1646387				
REFERENCE	2 (bases 1 to 2998)				
AUTHORS	Miyakawa, T.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-OCT-1990) Miyakawa T., Department of Fermentation Technology, Faculty of Engineering, Hiroshima University, Saijo, Higashi-Hiroshima 724, Japan				
COMMENT	See X54963 for related sequence.				
FEATURES	Location/Qualifiers				
source	1..2998				
	/organism="Saccharomyces cerevisiae"				
	/strain="A364A"				
	/db_xref="taxon:4932"				
	/clone_lib="lambda gt11"				
misc_feature	539..545				
	/note="consensus alpha-factor response element"				
misc_feature	663..669				
	/note="consensus alpha-factor response element"				
misc_feature	807..813				
	/note="consensus alpha-factor response element"				
gene	907..2721				
	/gene="CMP2"				
CDS	907..2721				
	/gene="CMP2"				
	/codon_start=1				
	/product="calmodulin-binding protein 2"				
	/protein_id="CAA38712.1"				
	/db_xref="GI:3560"				
	/db_xref="SWISS-PROT:P14747"				
	/translation="MSSDAIRNTEQINAAIKIENKTERPOSSSTPIIDSKASTVAAN STATSDLTQYLDGGRVSTNRRIMNKVPATISHVPTDEDVFPQNPGRHEFLRD HFKEGKLSAAQAAIRIVTLATELSEKPNLISVPAPITVCGDIHGQYFDLKLFEVGG DPATSYLFLGDYVDRGSFSECLTYLYSLKLNDFHLLRGNHECKHLTSYFTFKN EMLKYNLDIYEKCESFNPLAALMNGQVLCVHGGISPELNSLODINNLRFEIP SHGLMDLLWADPIEYDEVDLKDTEEDIYNSKTMVPHGKMARSDMFVNSVRGC SYAFTYRAACHFLQETGLSTIRAEADQAGTMYKNTKTLGFPFSLTLFSAPNYLDIT YNNKAAILYNNWNINRFQNNTPHYLPDPMDFVTSPLPFGVEKVTMLVAILNIC TEDELDNPTVEELVGDYDKLPQAGKSEATPQATSPASPKHASITLDDDEHRRKALRNK ILAVAKVRMYSVLREETNKVQFLKDHNSGVLPRGALSNGVKGLDEALSTFERARKHD LINEKLPSLDLKENKNKYKVKQKVHEHDAKNDSK"				
misc_binding	2407..2475				
	/gene="CMP2"				
	/bound_moiety="calmodulin"				
BASE COUNT	878 a 673 c 558 g 889 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	1.36e-98	Length:	2998		
Score:	104.00	Matches:	104		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	8	Gaps:	0		
US-09-814-661a-2 (1-104) x SCCMP2 (1-2998)					

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 110 ATGCAGAAATCCCAAGACTACTTTTACCGTCAAAATCGTCCCAACAACAACAGCCCT 169
 QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPheMetAla 40
 170 TCCACATTGCGTACCGTACCATGCCGAAATTTAGAGGGTGCCCTTCCACCTATGGCT 229
 QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
 230 GAGGTTCCTATGTTGCTACTCAAAATCCATGGCAGCTCGCTTCTCCCTCCGCTTCT 289
 QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
 290 TCATTAGAAATGTGGGAAAGGATTTGGAGGAGACACTCAACTCTATCGATCATGACATG 349
 QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlnLysValGlu 100
 350 AACACAAACAAATTTGTTCTGGCGCACTAAATCTATGTTCAACACAGGTAAGTTCGAG 409
 QY 101 GluMetAspPhe 104
 410 GAAATGGAGCTTC 421

RESULT 2

SC9958

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

SOURCE

1..20951

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 9958"

complement(1..507)

/gene="PIF1"

complement(<1..507)

/note="YN9958.01c, incomplete PIF1 gene, len: 169, CAI:

Notes:

All CDS over 100 codons have been analysed. CDS that are completely

overlapped and those that are overlapped by more than 50% of their

length by a larger CDS have been omitted from this analysis.

Details of the omitted CDS are available on request. The more

significant matches with motifs in the PROSITE database are also

included but some of these may be fortuitous. The length in

codons/amino-acids is given for each CDS as is the calculated codon

adaptation index (CAI)

Cosmid 9958 is overlapped at the start of this sequence by cosmid

9745

and at the end of this sequence by cosmid 9827.

Bases 7395 to 7399 and 14397 to 14398 are ambiguous.

0.14"
/codon_start=1
/product="Piflp"
/protein_id="CAA86714.1"
/db_xref="GI:577135"
/db_xref="SWISS-PROT:P07271"
/translation="MPKWRSTLNHIIPRRPFCISFNSFLLKNVSHAKLSFMSRRG
FRSNP IQAQLKHPISLKSDDLDSDDEEPCIOLETEKQEKIITDIHKEDPV
DKPMRDKVMNFINKDSPLSWDMFKPIIQPOLISENSFQSSQKRSRSTGFKNP
LRPAKKES"
846..1976
/note="YM9958.02, unknown orf, len: 376, CAI: 0.15"
/codon_start=1
/product="unknown"
/protein_id="CAA86715.1"
/db_xref="GI:577136"
/db_xref="SWISS-PROT:P53397"
/translation="MSYKFGKLAINKSELCLANVLAQGSFRWIWDEKINQVSTWTKI
GOQKTSVILRODENEIEFVAVGDCGNQDAKTHMKYFRLDVSLKLFEDNVWIP
SKAFALSPQGIIRILAQEPWETLISFCISNNNISRIITMCSNLSGNTLITTDG
VAYHSPTSEELTSTRATEAKRLGFGYRAKYIETARKLVNDKASANTISDTYTLQS
ICKDAQYEDVRHLSYNGVGPVADVCVCLMGLHMDGIIVPDVHVSRIRAKRDYQISAN
KHLKELRTKYNALPSRKINLELDHILMLFKWGSYAGWAQQVLFKEIGTSGS
TTTGTTKKKMDIKTEALIVYQMKLKVLSLDLHKEAKID"
complement(2194..7233)
/note="YM9958.03c, similar to unknown C. elegans gene
YOL4_CAEEL_ZK370.4 CE00396, and E.coli YCHK SW:YCHK_ECOLI
P37053 and Shigella SW:YCHK_SHIFL_P37054; len: 1679, CAI:
0.17"
/codon_start=1
/product="unknown"
/protein_id="CAA86716.1"
/db_xref="GI:577137"
/db_xref="SWISS-PROT:Q04958"
/translation="MPSMCTTNTNTGNTKNTKNSIGSFNSNYSYRFOCLTDQI
ISEAOTLNSLNFVSWVSVFVPGASRMIFRYGWLATLSLLRIPKWIFFKLHHVQF
TISFWLILALAVIVFYTIMKERILSQKRLTPEFLPLENGTSGSSANINAASTQ
SANAPALGSTGASSIIDSXKLSLKDNEFETLSSYLDQDLSAIFKLYGERKPVF
HDTKNMKQMDGILLIDSTIGFAIVVEGTLOLYHEVDSDKHGDETDHSDTDG
LDQDREDEEDDDIDNVDYKSCSNLIDEDSVGYIHLKNGLGNFQLNLTVPKG
NPLTSLVSLNLTTHSMSSVGNFSELSSPIDTTSVNNMPCSEQNFENSTDSMTN
STNSFTFPSSMPKLVARATDCTIGLIIPQSFALTKAPRSASHIIQWLTKLHV
TFOTADYLGTLKMDIEVLLNKSIVIELPYLKEAVIRKFKTVDKSGSADLEPKP
KNASNSKPKPKAPSDGIQSLKIANANANTSSLSLKEPEFTHPPSSRRVLGS
RDFNPGDLSNPLSTMDILSPNPIHNNRNKSGINTSTNQHKRSRRSSNNAS
VHSKFSSLSPELRNAQLSTPLSDNTSVHDHIHPSVHLKGRVSPRNLLPTTFSF
AAQETEDSALRMALVEAMTLVGVKNMSVSSSIANMSLSNQLNEMYSRRPSN
ASFLMSPCTPSDISVASSFASPTQOTMLRILPKYITISNKRHNKSQDKKKPRAY
KEBLTNLDEFEDVKDPAGIQLKFKKGTTVIQNARGGLFYIISGKVSFNIVAKS
SVVSMKPEQVSAQSHKGENPHHTQHLIYSVSGGIYGVYSSLIGYSFVNIVAKS
DVYVGLSATLERLDKYFLIYRISDLSLKLSSRLKLHLALWHLRASSETLFS
QGDANGIYVNLGRQLQQQSLNSNTSEEVETQIITLGAQGESFGEVELTA
MNRYSITIVARDESLARIPRTELLALEHPSIMIRSVLAKKIIVGDRVPAALTGDP
LSIKENDFTSLIPPTKASYSSLSHKPQNTISCTIFRITILPISGLPVEAFAKL
VQAFQVGRITIGLQRTLTHGRHAFRLSKLQSGYFAELEEMFQTVVYISDTPV
KSNWRTAQGICILLADARSPSAEIEYEKLLNSKTARTLELILHPERIVEFP
LTHKLRYPWVSHHHIQFSLTGTTLMEGKMHVNLNGALMDKLIQTEFSRKTQQ
NISKLPLDSIKNTVNSHFSSRFMSKRYVTPVHRHKNLDFRLARILSGOAILGLVGG
GARGISHLGQIAIEEQIIPVDVIGTSIGSVFGVLYAKDYDLVPIYGRVKFAGRIS
STWRMLTDLTPVTSYTTGHEFNRIWTKTGEDRIEDFQIYQYCNSTNITDSVQEIHS
FGYAWRIYRASHSLAGLLPLENGSHLDGGYVDNLPTVMRARGCQITFAVDVGS
DDTPMEYGDLSNGFWIIFNRWNPFSHPNIPMAETQVRLGVYASNALEAKNTPG
VVYVRPIEETALDFSKFEETIYHGVVDYGRIFLOGLIIDDKMPYIPGSGEQLTNSQV
PEFLHHRNSI"
7395..7399
/note="ambiguous sequence, agaag or ggaag"
8358..8672
/note="YM9958.04, unknown orf, len: 104, CAI: 0.32"
/codon_start=1
/product="unknown"
/protein_id="CAA86717.1"
/db_xref="GI:577138"
/db_xref="SWISS-PROT:Q04964"

misc_feature

CDS

/translation="MONSODYFYAQRNCOQOQAPSTLRTVTMAEFRVPLPPMAEYPM
LSTQNSMGSSASASASLEWKEKLEERLNSIDHDMNNKFGSGELKSMFNOGKVEEM
DP"
9155..10969
/gene="CNA2"
9155..10969
/gene="CNA2"
/note="YM9958.05, CNA2 or CMP2 gene, protein phosphatase,
P2B2_YEAST_P14747, len: 604, CAI: 0.16"
/codon_start=1
/product="calcineurin"
/protein_id="CAA86718.1"
/db_xref="GI:577139"
/db_xref="SWISS-PROT:P14747"
/translation="MSSDAIRNTEQINAAIKIENKTERPOSSTTTPIDSKASTVAAN
STATETSRDLTOYTLDDGRVSVNRRNIMNKVPAITSHVPTDEELFGPNIPRHEFLRD
HKRREGKLSAQQAARIVTLATLFLSKPENLISVPAPITVCGDIIHGQIFOLLKLFYVGG
DPATTSYLFSDYVDRGFSFECLYILYSLKLNFDHFWLLRGNHCKHLTSTYTFKN
EMLKLDLYIEKCESFNPLAALMGQYLCVHGGI SPENLSQDINNLRFRIP
SHGLMCDLWADPIEEDVDLKDTEEDIVNSKTMVPHHGMAPSRDMFVPSNRGIC
SYAFTYRAACHFLOETGLLSII RAHEAODAGYRMYNKNTLTGFPSSLTLTFSAPNYLDT
YANKAAILAYENNVMNIROFNTPHPHWPDPDMVTWSLPFVGEKVTMLVALNITC
TEDELNDTPVIEELVGTDKKLPQAKSEATPOPATISASPKHASILDDERRKALRNK
ILAVAKSRMYSVLRNENKRYKQFLKDHNSGVLPRGALNSGVKGLDEALSTFERARKHD
LINEKLPPSLDELKLNKKYKWKVQKVEHDAKDSK"
9755..9772
/gene="CNA2"
/note="P500125 Serine/threonine specific protein
phosphatases signature"
complement(111169..12283,12692..13151))
/note="YM9958.06c, orf len: 524, CAI: 0.48, similar to
inosine-5'-monophosphate dehydrogenases"
/codon_start=1
/product="putative inosine-5'-monophosphate dehydrogenase"
/protein_id="CAA86719.1"
/db_xref="GI:577140"
/db_xref="SWISS-PROT:P50094"
/translation="MSAAPLDYKKALEHLKTYSKDGLSVQELMDSTTRGGLTYNDFL
VLPLGNVTFSSAVSLQTKLTKITLTPVPSSPMDVTVEADMAIYMALGGIGFIHNN
CTPEQASAVKVKVPENGFINSPVISPITTVGEVKVMKRFEGSFVTEGDKCPG
KLVLGTSRDIOQLDEDSLVSVMYFNKPTGIGITLKEGNEILLKQTKGKGLIIVDD
NGNLVSMRADLNMKNONPLASKSATTKPQCAIGITTEADKERLRLVLEAGLDVV
LIDSSQNSVQLNMKIWKTEPDLIEITAGNVATREQANLIAAGADGLRIGMSGSG
ICITQVMACGRPQGTAVNVQCFANQGVPCMDAGGVQNIHITKALALGSSVMMG
GLMAGTSPGEGEYFKDKRLKAYRGMSIDAMQKNGKNASTSRVFSFSDSVLVAQ
GVSGAVVDGSKIKKFTIPYLVNGLQHSQDQIGCESLTSLENVQNGEVREFERTASAQL
EGGVHNLHSYKRLYN"
complement(11730..11768)
/note="P500487 IMP dehydrogenase/GMP reductase signature"
complement(12284..12286)
/note="splice acceptor sequence, cag"
complement(12396..12402)
/note="splice branch sequence tactaac"
complement(12686..12691)
/note="splice donor sequence gtatgt"
13765..14301
/note="YM9958.07, unknown orf, len: 178, CAI: 0.15"
/codon_start=1
/product="unknown"
/protein_id="CAA86720.1"
/db_xref="GI:577141"
/db_xref="SWISS-PROT:Q04969"
/translation="MSSAKPINYSIPPELNLQALDEALPSVFARLNYSYVALLDAKLY
IGYSIAVAGLSFLLKPERDOIVTYOKLLVGAYFVLSLLFWFSRFIEKGYVYVCK
RGCTKEIYVTKTFEKNELPLYLVELQKKGKNGSKELKAKLEVKNVFNESGYLQDA
YKWFSEQHNVLDTKNE"
14397..14398
/note="ambiguous sequence, ct or cc"
complement(14508..16283)
/gene="CYB2"
complement(14508..16283)
/gene="CYB2"
/note="YM9958.08c, CYB2 gene, CYB2_YEAST_P00175,

misc_feature

gene

CDS

cytochrome b2 precursor, len: 591, CAI: 0.21"

/codon_start=1

/product="cytochrome b2 precursor"

/protein_id="CAA86721.1"

/db_xref="GI:5771142"

/db_xref="SWISS-PROT:P00175"

/translation="MLKVKPLKIKSKNCAATLRASKTRRLNTRAYGSTVPKSKSFEQ

DSRRKTSQWALRVCAILAAATSSVAYLNWHNGQIDNEPKLDMNKKQISPAEVAKHKKP

Alignment Scores:

Pred. No.: 9.53e-98 Length: 20951
 Score: 104.00 Matches: 104
 Percent Similarity: 100.00 Conservative: 0
 Best Local Similarity: 100.00 Mismatches: 0
 Query Match: 100.00 Indels: 0
 DB: 8 Gaps: 0

US-09-814-661A-2 (1-104) x SC9958 (1-20951)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 Db 8358 ATGCAAAATTCACAGACTACTTTTACGCTCAAAATCGCTGCCAACACACAGCCCT 8417
 QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuProMetAla 40
 Db 8418 TCCACATTGCGTACCGTACCATGCGGAATTTAGAGGGTGCCTTTGCCACCTATGGCT 8477
 QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
 Db 8478 GAGGTTCCTATGTTCTACTCAAACTCCATGGCAGCTCCGCTTCTGCCCTCCCTCT 8537
 QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
 Db 8538 TCATTAGAATCTGGAAAGAGATTGGAGGAGACACTCACTCTATCATCATGACATG 8597
 QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
 Db 8598 AACACACAAATTTGGTTCTGCGCAACTAAAAATCTATGTTCAACACAGGTAAGGTGCGAG 8657
 QY 101 GluMetAspPhe 104
 Db 8658 GAAATGGAGCTTC 8669

RESULT 3

ACH251521 3089 bp DNA linear PLN 15-MAY-2001
 LOCUS Acronium chrysogenum pacc gene for putative transcription factor,
 DEFINITION exons 1-4.

ACCESSION AJ251521

VERSION AJ251521.1 GI:14132802

KEYWORDS pacc gene; transcription factor.

SOURCE Acronium chrysogenum.

ORGANISM Acronium chrysogenum

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acronium.

Schmitt,E.K., Kempken,R. and Kuck,U.

Functional analysis of promoter sequences of cephalosporin C

biosynthesis genes from Acronium chrysogenum: specific

DNA-protein interactions and characterization of the transcription

factor PACC

Mol. Genet. Genomics 265 (3), 508-518 (2001)

21298460

11405634

2 (bases 1 to 3089)

Schmitt,E.K.

Direct Submission

Submitted (08-DEC-1999) Schmitt E.K., Lehrstuhl fuer Allgemaine

Botanik, Ruhr-Universitaet Bochum, Universitaetsstrasse 150, 44780

Bochum, GERMANY

Location/Qualifiers

1..3089

/organism="Acronium chrysogenum"

FEATURES

source

/evidence=experimental

BASE COUNT

625 a 1123 c 754 g 587 t

gene /db_xref="taxon:5044"
 join(175..238,283..388,392..397,574..578,794..1006,
 1007..1078,1079..1282,1283..1358,1359..1446,1447..1504,
 1505..2865,3089)
 /gene="pacc"
 CAAT_signal 175..179
 CAAT_signal 206..210
 enhancer /gene="pacc"
 223..228
 /gene="pacc"
 /note="putative PACC binding site"
 enhancer 233..238
 /gene="pacc"
 /note="putative PACC binding site"
 gene complement(251..256)
 enhancer /gene="pacc"
 complement(251..256)
 /gene="pacc"
 enhancer 283..388
 /gene="pacc"
 /note="putative PACC binding site"
 enhancer 392..397
 /gene="pacc"
 /note="putative PACC binding site"
 TATA_signal 574..578
 /gene="pacc"
 join(794..1006,1079..1282,1359..1446,1505..2865)
 /gene="pacc"
 join(794..1006,1079..1282,1359..1446,1505..2865)
 /gene="pacc"
 /function="pH regulation"
 /codon_start=1
 /product="putative transcription factor"
 /protein_id="CAC38840.1"
 /db_xref="GI:14132803"
 /translation="MSPPAPDAQPSGSSSDNSNADSKSTTPPPSTTTTSAASIS
 AAVPDDNLTCRWACNOKFPNPTLYDHICERHVGKSTNNLNTCOMNSCRTTVK
 RDITSHIRVHVPLKPKCECGKSKRPQDLKKHVTHADDSVLYRPSODNHNGGM
 NYRHLNKPTSYDHNGMRTQYQHQPAPNAIYAPQPSNYGLYFNQPLNNHQT
 EHLGYAPPQYDRKRAFDMVDYDFSSAKREIDPSSYSQIGRSLPLUGLTSPSGP
 MPATEQHYVPOHAGIAHGAFTQNPQAQYLLPMPNARTQKDLVQIDNLLGQM
 DTIVENANHTAGVQIHGHGDIHGYRSGSPPIAQKSGPMGPGYHQPVSAAT
 MASPLTASSTGTTPAVTPPSSAVSVTSGHSPSPAPASPOSRSSTSCSVYPSLPT
 SLPAVSQCGOSTTTTLGPFSESRERYSGMQLQARAAPRSTDNHNHNCAS
 PKAPVPALASPSESVSYSEATREEQYDRWLENMRVIESLSREYVRCRLRREYVE
 EOENPHQCGGEAARKERDGNPMDVDARSFSPAREQPGRPREGSSLYPLRMPGD"

794..1006

/gene="pacc"

/number=1

1007..1078

/gene="pacc"

/number=1

1079..1282

/gene="pacc"

/number=2

1283..1358

/gene="pacc"

/number=2

1359..1446

/gene="pacc"

/number=3

1447..1504

/gene="pacc"

/number=3

1505..2865

/gene="pacc"

/number=4

3089

/gene="pacc"

/evidence=experimental

polyA_site


```
ANPTATREHVQAVRAAANDYEASDAFAETQRTPAASGRDAEPSVQWAFRVQVPVND
DRLVVGQDVRERQRLAEELRSEELHRLVNNMTDTVLVTDGDTGTHLVCPNNAHYIF
GYTAEETRAFQTDIDELGCDPVGDSLSRNGVTNVECTATDANGDGTHTLVNVRVNS
IQGTRTLYSCROVTRKQREDALAQLOETSDLLXACTEATATRVVDDGTAAALPAGG
IAYTRFDHENRLTPVAADNLAAYDADAGELRDDRSPVGTAFVADRVYDUGAGGEW
DAGVAAGLAEHTALPIGDHGVVLLAGTDRGRSDDALGAVGIEIARLLAATLBAALDRLD
FEALRERDQRLQONRRLSINQVNEIIRVEDGALVDADSRADIEAVCERLTAGDR
FAFVWGDATANDHVTTPRAWAGDASSYLDAAALPTASGSEPAVTAADRSVTLVP
NVADRRDADPAAALAHGFNSVPLVSDVDFLFTVTVYVDFDQSKLLGRLAADFQCLTVSEVAGST
LGQTVAGLSSAORKAALDGTTPVTVYAVFDQSKLLGRLAADFQCLTVSEVAGST
GQTRVYVSDGADFDVAVAGATALVDVIGDVQVATGDSGVVSLRGLRSLRALAQ
GVTRRQLTATPESGTLVAAAPDAATAARAADSLADAFASVDVTAARRRQSSADDAQGP
SLTRQVEQAVQAYHSGFFDPSHAVSGAVVAETLGISNTAFYDHVNRIESKVFDAVED
DGGCSWVE"
7901..9190
/gene="gdh"
/feature="VNG0161G"
7901..9190
/gene="gdh"
/feature="gdh"
/feature="gdh"
/codon_start=1
/transl_table=1
/product="glutamate dehydrogenase"
/protein_id="AAG18779.1"
/db_xref="GI:10579809"
/transl_table=1
/translation="MAQTTPPESSAPDSEPTAETARROLOAAAEHDIADGVIER
LTHTRVQVSVVPERDDGTTVYDGYRAQDDVYGPVKGGLRVHPGSAEACVGLSM
WMTWKAAMDLPFGAGKGVVDPKTLISADEHERLTRFAELRDEQPSQDIPAPDM
GTDQATWAFNDAYSQOGETVPGVYTKGPPVAGSGHGRAPRAEGRSVAIATREALINY
DIPDQATWAFVGGTGVGANAALLDDQARVAVSDVNGGVLDGDLTHAIPSHGN
QPAAMRHADPATLNEELLEDDVDVPIPAAGVNIITAAANDRIQADIVVEGANGPTT
SAADRIIEERAVPTPIDLANAGVTYSYFELQDINRRTWSPERVDRLESEMSLAW
NAVSEVDDGSLNRDAAYVVALQRIQRAKEARGLWP"
9073..10875
/gene="alk"
/feature="VNG0162G"
9073..10875
/gene="alk"
/feature="alk"
/feature="alk"
/codon_start=1
/transl_table=1
/product="medium-chain acyl-CoA ligase"
/protein_id="AAG18780.1"
/db_xref="GI:10579810"
/transl_table=1
/translation="MERRAQGRRRRPELAGRLCRAATHRPCEGSPWALALICTAL
GVPTPTTITDVCDDGGMHRLPLVTFIDRARTYGDTEAVVATTGRRVYSELADR
ADRTSAVLAAGIDSDRVAVLDPNTHLEAAYGAMOLGAHTPLAVRLTPDDYAVI
LADAGADAIYADAAVADREAVRDOIPTTFEITDDRAVGDQDADALLADPDYX
DRPDMAEDDVITINYSCTGDPKGVCRTHNEVVHALLSTHTITDITDITLWLPX
FHANGWHIFAITGAGATHVCTRGVDAASVFDITQAEAVSYLCAPIVNLMLDYHEA
HDVAATGDADVRVATAGSAPPEATIRAVESALDWSLMHYGATETGLIATSDADRL
AAGGSRAMKRGKIGLGTTELTVDEHGNVPRDNETIGEIVVSGNQVMDRYWKKP
DATERATFTRVDGYFHMGDFAVDOTGTVSLDRKKDIIISGGENISIELEDFEH
DAGEVAVPSPDHDMKETPKAFVVPASGDPNPGVTITLVEYTRAEASFRTVRRV
EFVDELPTATGKVKQKVELRAAEFADEESLVGGQ"
complement(10891..13506)
/gene="mutS1"
/feature="VNG0163G"
/feature="VNG0163G"
complement(10891..13506)
/gene="mutS1"
/feature="mutS1"
/feature="mutS1"
/codon_start=1
/transl_table=1
/product="mismatch repair protein"
/protein_id="AAG18781.1"
/db_xref="GI:10579811"
/transl_table=1
/translation="MGIVDEFOALKAETDADLLAMQVGFYFEFFAADARTVASVLDLQ
VSEKNSGSSTPMAGVVPVDDLTPLYLAALVERGYKVAVAESETDAGDIERIERVTP
GTLAATDADPRYLAAYVREAGDGLAFVDVITGQFRTVRGADRAVATLEYLFPAP
EVLPCALGRDDDFLGLVRLRTDALTLDHAGAFDAGRATHRVRFQFQCDGVIESLGA
ADGPVRAAGAAGVYIAADEGVLAYSRITQPPFGGDDHVELDATTORNLLETMTG
SDGSLLATIDHTSAAGRLAAWVTRPTRDRARELDRRQAVGALADAAALARDALGDV
```

```
Alignment Scores:
Pred. No.: 7.71 Length: 13571
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 1 Gaps: 0

US-09-814-661a-2 (1-104) x AE004982 (1-13571)

Qy 52 GlySerSerAlaSerAlaSerAlaSer 61
|||||
Db 12269 GGATCGTCCGCCAGTGGCTCGCAAGCTCQ 12298

RESULT 5
AC020305 19502 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
DEFINITION pieces.
ACCESSION AC020305
VERSION AC020305.1 GI:6664592
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 19502)
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Ephyrdoidea; Drosophyllidae; Drosophila.
TITLE Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
JOURNAL Adams.M. and Venter.J.C.
COMMENT Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212709 by the submitter.
For more information on this record e-mail to fly@celera.com..
* NOTE: this is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source 1..19502
/organism="Drosophila melanogaster"
BASE COUNT 5273 a 3911 c 4269 g 6049 t
ORIGIN

Alignment Scores:
Pred. No.: 11.1 Length: 19502
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 2 Gaps: 0

US-09-814-661a-2 (1-104) x AC020305 (1-19502)

Qy 53 SerSerAlaSerAlaSerAlaSerSerLeu 62
|||||
Db 7303 TCCTCCGCTCGGCCGTCGGCATCCTCTTA 7332

RESULT 6
AC109752/c 64095 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-10324, *** SEQUENCING IN PROGRESS
DEFINITION ***; 13 unordered pieces.
ACCESSION AC109752
VERSION AC109752.3 GI:21738281
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 64095)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
```


AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisa, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogwu, M., Okwuonu, G., Ogunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
2 (bases 1 to 64095)
Worley, K.C.

Direct Submission

Submitted (07-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 64095)
Worley, K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 12, 2002 this sequence version replaced gi:18860267.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GQUB

Center clone name: CH230-10B24

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 50906 bases at least Q40

Consensus quality: 51674 bases at least Q30

Consensus quality: 52174 bases at least Q20

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html);

* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 1583: contig of 1583 bp in length
* 1584 1683: gap of unknown length
* 1684 3277: contig of 1594 bp in length
* 3278 3378: gap of unknown length
* 3379 4575: contig of 1198 bp in length
* 4576 4675: gap of unknown length
* 4676 6535: contig of 1860 bp in length
* 6536 6636: gap of unknown length
* 6637 9287: contig of 2652 bp in length
* 9288 9388: gap of unknown length
* 9389 13029: contig of 3642 bp in length
* 13030 13129: gap of unknown length
* 13130 16992: contig of 3863 bp in length
* 16993 17092: gap of unknown length
* 17093 20685: contig of 3593 bp in length
* 20686 20785: gap of unknown length
* 20786 25465: contig of 4680 bp in length
* 25466 25565: gap of unknown length
* 25566 29815: contig of 4250 bp in length
* 29816 29915: gap of unknown length
* 29916 37399: contig of 7484 bp in length
* 37400 37499: gap of unknown length
* 37500 49640: contig of 12141 bp in length
* 49641 49740: gap of unknown length
* 49741 64095: contig of 14355 bp in length.

FEATURES

source

1..64095
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-10B24"

BASE COUNT 18160 a 11023 c 11424 g 17934 t 5554 others
ORIGIN

Alignment Scores:

Pred. No.: 36.5 Length: 64095
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-2 (1-104) x AC109752 (1-64095)

Qy 50 SerMetGlySerSerAlaSerAlaSerAla 59
|||||

Db 37058 TCAATGGGTTCAATCTGCTAGTCGTCAGCT 37029
|||||

RESULT 7

AC122111

LOCUS

DEFINITION

AC122111

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AC122111 153328 bp DNA linear HTG 29-MAY-2002
Rattus norvegicus chromosome 1 clone RP32-587C22, *** SEQUENCING IN PROGRESS ***, 54 unordered pieces.

AC122111.1 GI:21070651

HTG: HTGS_PHASE1

Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 153328)

Taenzer, S., Monti, J., Gloeckner, G., Goesele, C., Baumgart, C.,

Huebner, N. and Platzer, M.

Rat chromosome 1 genomic sequence

Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

2 (bases 1 to 153328)
Taenzer, S. and Platzer, M.
Direct Submission
Submitted (22-MAY-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany

3 (bases 1 to 153328)
Taenzer, S. and Platzer, M.
Direct Submission
Submitted (29-MAY-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany

----- Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: <http://genome.imb-jena.de/>
Contact: gscj-submit@genome.imb-jena.de
----- Project Information

Center project name: RA31
Center clone name: RP32-587C22

----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 114285 bases at least Q40
Consensus quality: 125941 bases at least Q30
Consensus quality: 134029 bases at least Q20
Quality coverage: 3.0

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1007: contig of 1007 bp in length
1008 1107: gap of unknown length
1108 2277: contig of 1170 bp in length
2278 2377: gap of unknown length
2378 3464: contig of 1087 bp in length
3465 3564: gap of unknown length
3565 4802: contig of 1238 bp in length
4803 4902: gap of unknown length
4903 6163: contig of 1261 bp in length
6164 6263: gap of unknown length
6264 7422: contig of 1159 bp in length
7423 7522: gap of unknown length
7523 8554: contig of 1032 bp in length
8555 8654: gap of unknown length
8655 9816: contig of 1162 bp in length
9817 9917: gap of unknown length
9918 11705: contig of 1789 bp in length
11706 11805: gap of unknown length
11806 13263: contig of 1458 bp in length
13264 14817: contig of 1454 bp in length
14818 14917: gap of unknown length
14919 16027: contig of 1110 bp in length
16028 16127: gap of unknown length
16128 17776: contig of 1649 bp in length
17777 17876: gap of unknown length
17877 19426: contig of 1550 bp in length
19427 19526: gap of unknown length

19527 20782: contig of 1256 bp in length
20783 20882: gap of unknown length
20883 22082: contig of 1200 bp in length
22083 22182: gap of unknown length
22183 23399: contig of 1217 bp in length
23400 23499: gap of unknown length
23500 25304: contig of 1805 bp in length
25305 25404: gap of unknown length
25405 27312: contig of 1908 bp in length
27313 27412: gap of unknown length
27413 28442: contig of 1030 bp in length
28443 28542: gap of unknown length
28543 29816: contig of 1274 bp in length
29817 29916: gap of unknown length
29917 31089: contig of 1173 bp in length
31090 31189: gap of unknown length
31190 32801: contig of 1612 bp in length
32802 34699: contig of 1798 bp in length
34700 34799: gap of unknown length
34800 36596: contig of 1797 bp in length
36597 36697: gap of unknown length
36698 39198: contig of 2501 bp in length
39199 40672: contig of 1375 bp in length
40673 40772: gap of unknown length
40773 41946: contig of 1174 bp in length
41947 42046: gap of unknown length
42047 44500: contig of 2454 bp in length
44501 44600: gap of unknown length
44601 46269: contig of 1669 bp in length
46270 46369: gap of unknown length
46370 48382: contig of 2013 bp in length
48383 48482: gap of unknown length
48483 50737: contig of 2255 bp in length
50738 50837: gap of unknown length
50838 53072: contig of 2235 bp in length
53073 53172: gap of unknown length
53173 55286: contig of 2114 bp in length
55287 55386: gap of unknown length
55387 56331: contig of 1245 bp in length
56332 59256: contig of 2525 bp in length
59257 59356: gap of unknown length
59357 62487: contig of 3131 bp in length
62488 62587: gap of unknown length
62588 64913: contig of 2326 bp in length
64914 65013: gap of unknown length
65014 67876: contig of 2863 bp in length
67877 67976: gap of unknown length
67977 70689: contig of 2713 bp in length
70690 70789: gap of unknown length
70790 74173: contig of 3384 bp in length
74174 74273: gap of unknown length
74274 76795: contig of 2522 bp in length
76796 76895: gap of unknown length
76896 81600: contig of 4705 bp in length
81601 81700: gap of unknown length
81701 85208: contig of 3508 bp in length
85209 85308: gap of unknown length
85309 90473: contig of 5165 bp in length
90474 90573: gap of unknown length
90574 94867: contig of 4294 bp in length
94868 94967: gap of unknown length
94968 99717: contig of 4750 bp in length
99718 99817: gap of unknown length
99818 104267: contig of 4450 bp in length
104268 104367: gap of unknown length
104368 112186: contig of 7819 bp in length
112187 112286: gap of unknown length
112287 115970: contig of 3684 bp in length
115971 116070: gap of unknown length
116071 123436: contig of 7366 bp in length

```
* 123437 123536: gap of unknown length
* 123537 132458: contig of 8922 bp in length
* 132459 132558: gap of unknown length
* 132559 141339: contig of 8781 bp in length
* 141340 141439: gap of unknown length
* 141440 153328: contig of 11889 bp in length.
```

FEATURES

source

```
1. 153328
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/chromosome="11"
/clone="RP32-587C22"
/clone_lib="RPC1-32 BAC Library"
43517 a 32391 c 30122 g 41998 t 5300 others
```

BASE COUNT

ORIGIN

```
Alignment Scores:
Pred. No.: 87.4 Length: 153328
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 2 Gaps: 0
```

```
US-09-814-661A-2 (1-104) x AC122111 (1-153328)
```

```
Qy 53 SerSerAlaSerAlaSerAlaSerLeu 62
Db 74548 TCTTCGCTCTGCTCTGCTCTCTCTCTC 74577
```

RESULT 8

AC027026/c

LOCUS

```
Homo sapiens chromosome 11, clone RP11-589F4, linear. PRI 28-JUN-2002
```

ACCESSION

AC027026

VERSION

AC027026.10

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 155376)

REFERENCE

Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Boguslavsky, L., Bouckghalter, B., Brown, A., Burkett, G.,

Campopiano, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Graland-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 155376)

REFERENCE

Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckghalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Dwyer, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 28, 2002 this sequence version replaced gi:15144544.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L8830

Center clone name: 589_F_4

Location/Qualifiers

1. 155376

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="RP11-589F4"

/clone_lib="RPC1-11 Human Male BAC"

9. 196

/rpt_family="L1R33"

complement(1117..1494)

/rpt_family="MSTA"

complement(1697..2184)

/rpt_family="L2"

2116..2125

/note="x30 qual SNGL region"

2575..2632

/rpt_family="AT-rich"

3064..3344

/rpt_family="MER54A"

complement(3345..3646)

/rpt_family="AluSx"

3647..3848

/rpt_family="MER54A"

3873..4190

/rpt_family="L1MA4"

complement(5283..5597)

/rpt_family="L1P1"

5907..6459

/rpt_family="L1MC2"

complement(6460..7728)

/rpt_family="L1P5"

7729..9245

/rpt_family="L1MC2"

9246..9556

/rpt_family="AluSx"

9557..10318

```

repeat_region /rpt_family="LIMC2"
10361..10614
/rpt_family="MER74B"
10621..11797
/rpt_family="LIMC2"
11818..12141
/rpt_family="AluJb"
12156..12412
/rpt_family="LIM3d"
12388..12591
/rpt_family="LIM3de"
12602..12797
/rpt_family="LIM3de"
12818..12906
/rpt_family="LIM3de"
13011..13111
/rpt_family="LIM3de"
13121..13220
/rpt_family="LIM3de"
13259..14254
/rpt_family="LIM3de"
14258..14497
/rpt_family="LIM3de"
14509..18461
/rpt_family="LIPA10"
18462..18485
/rpt_family="GAAAA)n"
18486..20925
/rpt_family="LIPA10"
20976..21515
/rpt_family="LIMA6"
21529..22435
/rpt_family="LIMA6"
22436..22620
/rpt_family="AluSp"
22621..24411
/rpt_family="LIMA6"
24412..24639
/rpt_family="LIMA6"
24640..25555
/rpt_family="LIMA6"
25556..27115
/rpt_family="LIPA14"
27116..27510
/rpt_family="LIPA14"
27511..27535
/rpt_family="LIMA6"
27536..27828
/rpt_family="AluSx"
27829..27843
/rpt_family="LIMA6"
complement(27844..28104)
/rpt_family="LIPA10"
28105..28261
/rpt_family="LIMA6"
28329..29211
/rpt_family="LIME2"
complement(31233..31412)
/rpt_family="MIR"
31763..31870
/rpt_family="L2"
31882..31912
/rpt_family="(CATG)n"
complement(32526..32720)
/rpt_family="MIR"
complement(33209..33510)
/rpt_family="AluY"
complement(33622..33769)
/rpt_family="MIR3"
34649..35041
/rpt_family="LTR33"
36230..36254
/rpt_family="(TTCA)n"

```

```

repeat_region complement(36259..36358)
/rpt_family="L2"
repeat_region complement(36692..36954)
/rpt_family="LIMEC"
37118..37309
/rpt_family="MIR"
repeat_region complement(37321..37569)
/rpt_family="LIMEC"
repeat_region complement(37561..41672)
/rpt_family="LIPA7"
repeat_region complement(41641..42058)
/rpt_family="LIMEC"
repeat_region 42061..42108

Alignment Scores:
Pred. No.: 88.6 Length: 155376
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: Gaps: 0

US-09-814-661A-2 (1-104) x AC027026 (1-155376)
QY 53 SerSerAlaSerAlaSerAlaSerSerLeu 62
Db 84824 TCCTCTGCCTCTGCCTCTGCCTCCAGCTTG 84795

RESULT 9
AC079887/c 168761 bp DNA linear PLN 29-AUG-2001
LOCUS Oryza sativa chromosome 3 BAC OSJNBa0040E01 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC079887 GI:13112227
VERSION AC079887.16
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 168761)
AUTHORS Buell,C.R., Yuan,Q., Ouyang,S., Moffat,K.S., Hill,J.N., Burr,P.C.,
Hsiao,C.J., Zismann,V., Pai,G., Bowman,C.L., Fujii,C.Y.,
VanAken,S.E., Bowman,C.L., Craven,B., Utterback,T.R., Khalak,H.,
Feldblyum,T.V., Quackenbush,J., White,O., Salzberg,S.L. and
Fraser,C.M.
TITLE Oryza sativa chromosome 3 BAC OSJNBa0040E01 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 168761)
AUTHORS Buell,R.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 168761)
AUTHORS Buell,R.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 4 (bases 1 to 168761)
AUTHORS Buell,R.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE 5 (bases 1 to 168761)
AUTHORS Buell,R.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT On Feb 23, 2001 this sequence version replaced g1:13096047.
Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0040E01 is from Oryza sativa chromosome 3

```

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan and Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), FgenesH (<http://www.softberry.com/>), and Gensplicer (Mihaila Pertea and Steven Salzberg, contact.mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

source

```

1. .168761
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosomes="3"
/map="R2404"
/clone="OSJNBa0040E01"
gene
1. .511
/gene="OSJNBa0040E01.9"
/note="similar to unknown protein CW14 GB:BAA87958
GI:6520233 (Arabidopsis thaliana)"
join(<1. .63,191. .313,398. >511)
/gene="OSJNBa0040E01.9"
join(<1. .63,191. .313,398. .511)
/gene="OSJNBa0040E01.9"
/codon_start=1
/product="unknown protein 5'-partial"
/protein_id="AAK16169.1"
/db_xref="GI:13236647"
/translocation="KLIVOTYQKPVLSRQHKFFKPNYFEIDLQVHRPFSIRKGLE
AFRELRKHGVLDTGIQAKAELEPHVLCRMLNKIDFADSGQITLIMSSE"
543. .590
/rpt_family="(GA)n"
1090. .1110
/rpt_family="(CAA)n"
complement(1253. .4472)
/gene="OSJNBa0040E01.10"
/note="similar to DNA binding protein 2 GB:AAD16139
GI:4322940 (Nicotiana tabacum)"
complement(join(<1253. .1493,2847. .2990,3128. .3786,
4197. >4472))
/gene="OSJNBa0040E01.10"
complement(join(1253. .1493,2847. .2990,3128. .3786,
4197. .4472))
/gene="OSJNBa0040E01.10"
/codon_start=1
/product="putative DNA-binding protein"
/protein_id="AAK16171.1"
/db_xref="GI:13236649"
/translocation="MADRRSDGGGMMQOQFTSPQERVFDDGGVPGQVAAAPYGSDF
DQSSYMAIIAAGVGVGQVPTAAPVAVEDVAAAPGISLAPQFSMAAAPPESYQ
HPATLVSPPLAAGLPYPPYLVGVDAVPQWPPRPPPSFVLDLAAAPAEQRHS
MOQLLRAAFGGMHAAAPAAAPAAAEIOAPKDGYNWRKYQKQKLDKAEPSRYK
CTRDCPVKKIYVERSDGCIKEITYKGRHSRPRVPRRGGAASSSSMAAGTDHNA
GAADAADAEDDESDDDDTLLHEDDDGEGDHGVDGVGVQVRVRPKPIILQTRS
EVDLLDDGYRWRYGQKVVKNRPRRSYKTCADGNCVRKQIRASADPKCVLITTYG
RHNDPPGPRPAAANLQMPGPAAMRLAGGTAAHQOPSGGAHQMKET"
complement(2132. .2180)
/rpt_family="AT-rich"
complement(3140. .3242)
/rpt_family="(CGA)n"

```

```

repeat_region
repeat_region
repeat_region
repeat_region
gene
mrna
CDS
repeat_region
repeat_region
repeat_region
gene
mrna
CDS
repeat_region
repeat_region
repeat_region
gene
mrna
CDS

```

```

3201. .3329
/rpt_family="(CGG)n"
complement(3500. .3561)
/rpt_family="(CGG)n"
3605. .3776
/rpt_family="(CGG)n"
complement(4424. .4488)
/rpt_family="(CGG)n"
complement(6280. .8524)
/gene="OSJNBa0040E01.11"
/note="predicted by fgenesH"
complement(join(<6280. .8570,6986. .7132,7219. .7405,
7494. .7612,7698. .7815,8030. .8124,8264. >8524))
/gene="OSJNBa0040E01.11"
complement(join(6280. .6570,6986. .7132,7219. .7405,
7494. .7612,7698. .7815,8030. .8124,8264. .8524))
/gene="OSJNBa0040E01.11"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK16173.1"
/db_xref="GI:13236651"
/translocation="MAEALVAVLRLAASAAATARPQSRSGHSCAARVPCPCPPFR
RGLCARAAVAGPPEVDDDDAMTIDNLRFFDVNKGNGAFYQFAHRVLOGISTR
LSVTIGEDDLISLLQSLIKQASSQISEVDEESEEWEYKIKETNMFTVDKIQQVG
FFQEEKAFALRYQTAGMLETVRAGVLGDEDTGEESPKNLKIPSRSPISVENCILYSR
EGNGRVAFHIMDKPGLVDMLIIFHEKQGEVPLMYSSDDADITNDRIAPLIGRWEG
RSVTKRSGVYGATSEADTVLLEKDRNGQLILDNMSTKSGSSTTTTIVHTGSANNL
LOFDGGYEMTLPLGGMYGYPDTDIGKIVNDMSFHLFCWMESPGRRQRLVRTYSAG
LAVSYTFETKV"
8202. .8241
/rpt_family="(GAA)n"
complement(8350. .8379)
/rpt_family="GC-rich"
complement(8792. .8870)
/rpt_family="AT-rich"
complement(9182. .11589)
/gene="OSJNBa0040E01.1"
/note="nearly identical to translation initiation factor
5A precursor protein GB:CA65463 GI:6689385 (Senecio
vernalis): EST AU057661, AU108424 from this gene"
complement(join(<9182. .9584,9676. .9750,10035. .10091,
10177. .10299,11282. .11431,11539. .11589))
/gene="OSJNBa0040E01.1"
complement(join(9480. .9584,9676. .9750,10035. .10091,
10177. .10299,11282. .11407))
/gene="OSJNBa0040E01.1"
/codon_start=1
/product="translation initiation factor 5A"
/protein_id="AAK16176.1"
/db_xref="GI:13236654"
/translocation="MSDSEHHFESKADAGASKTYPOQAGTIKNGHIVIKNRPKVV
EYSTKTKGHGAKCHFVAIDIFNGKKLEDIVPSSHCNDVPHVNRDYDOLIDISEGF
VSLLESGGTGKDDLRLPSDEALLTQIKDGFAGKCOLIVTMSAMGEEQICALKDIOFK
N"
complement(12250. .12294)
/rpt_family="AT-rich"
complement(17604. .17644)
/rpt_family="GC-rich"
complement(18787. .24766)
/gene="OSJNBa0040E01.12"
/note="predicted by fgenesH"
complement(join(<18787. .20771,21761. .22523,22653. .22887,
24660. >24766))
/gene="OSJNBa0040E01.12"
complement(join(18787. .20771,21761. .22523,22653. .22887,
24660. .24766))
/gene="OSJNBa0040E01.12"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK16179.1"
/db_xref="GI:13236657"
/translocation="MAVVVPARFLVRLMLRVLRNLNPLQGAASVVEVNOILARMGAKV
EGDNYMLGYAMGDFNMDANGRSYPYHDEKMSNGHMCNGFTKPNANGYSEYDEKMLTR

```

TMLEHAIHQVYELHRYVYKIQDLKMQYQNKDIYAYPMLEDAKTKTNSPQLPPNGA
KSWPIOTPPNSITVKKASIAEHGVNMHPLKFLREGSVQSSPNGFPDSDVNAARQGT
FDLQLSADHYVDONASQSDPTDLGVADPKKPNONADLTIVSPGELGRFSDNASTSG
LHATNNVCGRQVDNLNERITGTYGRANGSVRGLSTYLENSHQSIKLPSTANFYN
KEYSEKHLDDGTSNFFAANAQKQEEKQIDKQVSSIHVFTPRYSDANPQMSIK
GVDSRANNOFFHQGNGSIGWFAARPLEAPINFPRLDRSHNSHGALAPPMSIP
RIDHPSGASPTGCTVDPRSSAINNATQPIPSFKGSSVYNSQISGTLKVKKNEDLD
GNCGFALDPTCASRPOHVAISSDEQTECLMFHSAHRHNFHAFNDKPKNFNLN
EALSDQEDYLVEODGVSLSLPOSKASGFPWLKTKTDTCTPSPDLONPKVFAHSNR
EYIDLSNWDKREAAHTHLSDSASTSLDCGVKESQDCGKIKDEARCDTITREVA
CINTQESATCLPVLQCYVPGDDKAANGDKKSSAPVRNFIDLNDADPNEDNESSVV

Alignment Scores:

Pred. No.: 96.2 Length: 168761
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 8 Gaps: 0

US-09-814-661A-2 (1-104) x AC079887 (1-168761)

QY 53 SerSerAlaSerAlaSerAlaSerSerLeu 62

Db 55141 AGCTCCGCGCTCGCGCTCGCTCTCTCTC 55112

RESULT 10

AC111888/c

LOCUS Rattus norvegicus clone CH230-47K23, *** SEQUENCING IN PROGRESS
DEFINITION *** 78 unordered pieces.

AC111888 169179 bp DNA linear HTG 13-JUL-2002

AC111888.2 GI:21736654

HTG: HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

AUTHORS

1 (bases 1 to 169179)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okuwono,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleceyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zbrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 169179)
Worley,K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 169179)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701734.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu/
----- Project Information
Center project name: GOKT
Center clone name: CH230-47K23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 83033 bases at least Q40
Consensus quality: 88610 bases at least Q30
Consensus quality: 93298 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a working draft sequence. It currently
consists of 78 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1002: contig of 1002 bp in length
1102: gap of unknown length
1103 2746: contig of 1644 bp in length
2747 2846: gap of unknown length
2847 3954: contig of 1108 bp in length
3955 4055: gap of unknown length
4055 5134: contig of 1080 bp in length
5135 5234: gap of unknown length
5235 6346: contig of 1112 bp in length
6347 6446: gap of unknown length
6447 7580: contig of 1134 bp in length
7581 7680: gap of unknown length
7681 9186: contig of 1506 bp in length
9187 9286: gap of unknown length
9287 10410: contig of 1124 bp in length
10411 10510: gap of unknown length
10511 11723: contig of 1213 bp in length
11724 11823: gap of unknown length
11824 13410: contig of 1587 bp in length
13411 14539: contig of 1029 bp in length
14540 14639: gap of unknown length
14640 15684: contig of 1045 bp in length
15685 15784: gap of unknown length
15785 16944: contig of 1164 bp in length

* 16949 17048: gap of unknown length
 * 17049 18173: contig of 1125 bp in length
 * 18174 18273: gap of unknown length
 * 18274 19466: contig of 1193 bp in length
 * 19467 19566: gap of unknown length
 * 19567 21033: contig of 1487 bp in length
 * 21034 21133: gap of unknown length
 * 21134 22995: contig of 1842 bp in length
 * 22996 23096: gap of unknown length
 * 23097 24157: contig of 1062 bp in length
 * 24158 24257: gap of unknown length
 * 24258 25408: contig of 1151 bp in length
 * 25409 25508: gap of unknown length
 * 25509 26862: contig of 1354 bp in length
 * 26863 26962: gap of unknown length
 * 26963 28255: contig of 1293 bp in length
 * 28256 28356: gap of unknown length
 * 28357 29928: contig of 1573 bp in length
 * 29929 30028: gap of unknown length
 * 30029 31812: contig of 1784 bp in length
 * 31813 33301: contig of 1389 bp in length
 * 33302 33401: gap of unknown length
 * 33402 34666: contig of 1265 bp in length
 * 34667 34766: gap of unknown length
 * 34767 36278: contig of 1512 bp in length
 * 36279 36378: gap of unknown length
 * 36379 37733: contig of 1355 bp in length
 * 37734 37833: gap of unknown length
 * 37834 39165: contig of 1332 bp in length
 * 39166 39265: gap of unknown length
 * 39266 40959: contig of 1694 bp in length
 * 40960 41059: gap of unknown length
 * 41060 43021: contig of 1962 bp in length
 * 43022 43121: gap of unknown length
 * 43122 44271: contig of 1150 bp in length
 * 44272 44371: gap of unknown length
 * 44372 45763: contig of 1392 bp in length
 * 45764 45863: gap of unknown length
 * 45864 47668: contig of 1805 bp in length
 * 47669 47768: gap of unknown length
 * 47769 49042: contig of 1274 bp in length
 * 49043 49142: gap of unknown length
 * 49143 50656: contig of 1514 bp in length
 * 50657 50756: gap of unknown length
 * 50757 51833: contig of 1077 bp in length
 * 51834 51933: gap of unknown length
 * 51934 53044: contig of 1111 bp in length
 * 53045 53144: gap of unknown length
 * 53145 54852: contig of 1708 bp in length
 * 54853 54952: gap of unknown length
 * 54953 56449: contig of 1497 bp in length
 * 56450 56549: gap of unknown length
 * 56550 57832: contig of 1283 bp in length
 * 57833 57932: gap of unknown length
 * 57934 60096: contig of 2164 bp in length
 * 60097 60196: gap of unknown length
 * 60197 62288: contig of 2092 bp in length
 * 62289 62389: gap of unknown length
 * 62390 64015: contig of 1627 bp in length
 * 64016 65598: gap of unknown length
 * 65599 65698: gap of unknown length
 * 65700 67470: contig of 1772 bp in length
 * 67471 67570: gap of unknown length
 * 67571 69191: contig of 1621 bp in length
 * 69192 69291: gap of unknown length
 * 69292 70940: contig of 1649 bp in length
 * 70941 71040: gap of unknown length
 * 71041 72754: contig of 1714 bp in length
 * 72755 72854: gap of unknown length
 * 72855 74323: contig of 1469 bp in length
 * 74324 74423: gap of unknown length

* 74424 76022: contig of 1599 bp in length
 * 76023 76122: gap of unknown length
 * 76123 78172: contig of 2050 bp in length
 * 78173 78272: gap of unknown length
 * 78273 79608: contig of 1336 bp in length
 * 79609 79708: gap of unknown length
 * 79709 81121: contig of 1413 bp in length
 * 81122 81222: gap of unknown length
 * 81223 83049: contig of 1828 bp in length

Alignment Scores:

Pred. No.: 96.5 Length: 169179
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.62% Indels: 0
 DB: 2 Gaps: 0

US-09-814-661a-2 (1-104) x AC111888 (1-169179)

Qy 53 SerSerAlaSerAlaSerLeu 62

Db 148400 TCCTCTGCTCTGCTCTCTCTCTCTC 148371

RESULT 11

AC093435

LOCUS

DEFINITION

AC093435

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 171689)

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,

Gocayne,J.D., Tabor,P., Williamson,A., Homsy,F.H.,

Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,

Ayele,M., Scott,G.S., Worley,K.W., Amaratunga,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,

Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,

Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,

Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,

Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,

Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,

Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,

Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,

Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,

Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 171689)

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Garg,N., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,

Hollins,B., Homsy,F., Howard,S., Huber,J., Huiy,K.S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,

TITLE

JOURNAL

REFERENCE

AUTHORS

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisedge, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

TITLE
JOURNAL

REFERENCE

AUTHORS

3 (bases 1 to 171689)

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amarutunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Butler, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Homsli, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisedge, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

TITLE
JOURNAL

REFERENCE

AUTHORS

JOURNAL

Submitted (25-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 171689)
BCM-HGSC.
Direct Submission
Submitted (25-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

5 (bases 1 to 171689)
BCM-HGSC.

Direct Submission

Submitted (03-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Aug 3, 2002 this sequence version replaced g1:21553182.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot>

FEATURES

Source

Location/Qualifiers

1..171689

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="3L"

/clone="Rp98-7A5"

108272

/note="G/A"

/function="unsure"

117095

/function="low quality"

131419..131422

/function="low quality"

153210

/function="unresolved homopolymeric run"

BASE COUNT 49306 a 35683 c 35732 g 50968 t

ORIGIN

Alignment Scores:

Pred. No.:	97.9	Length:	171689
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	9.62%	Indels:	0
DB:	3	Gaps:	0

US-09-814-661A-2 (1-104) x AC093435 (1-171689)

Qy 53 SerSerAlaSerAlaSerAlaSerLeu 62

|||||

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dethorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kurushi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, S., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunou, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Slismon, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 179838)
BCM-HGSC

Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 23, 2002 this sequence version replaced gi:15143970.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for

a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

Location/Qualifiers

1..179838

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="3L"

/clone="RP98-7E7"

25972..25977

/function="low quality"

54788..54790

/function="low quality"

54804

/function="low quality"

174839..174840

/function="low quality"

174942..174947

/function="low quality"

BASE COUNT 52143 a 38739 c 38564 g 50392 t

ORIGIN

Alignment Scores:

Pred. No.: 103

Score: 10.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 9.62%

Indels: 3

Gaps: 0

Length: 179838

Matches: 10

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-814-661a-2 (1-104) x AC091220 (1-179838)

Qy 53 SerSerAlaSerAlaSerAlaSerSerLeu 62

Db 164985 TCCTCCGGCTCGCGTCGTCATCCTCTTA 164956

RESULT 14

AC114454

LOCUS

DEFINITION

AC114454

AC114454.2 GI:21745638

HTG: HTGS_PHASE1

Norway rat

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 186357)

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Albrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbieri, J., Benzon, J., Blinag, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.B., Dathorne, S.R., David, R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,

Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,

Hollins, B.,

Hollins, B.,

Hollins, B.,

Hollins, B.,

Hollins, B.,

Hollins, B.,

Hollins, B.,

Hollins, B.,

Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozato, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oregueta, N., Oviold, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 186357)
Worley, K.C.
Direct Submission
Submitted (09-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 186357)
Worley, K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:19310248.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSXH
Center clone name: CH230-72B12
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 131284 bases at least Q40
Consensus quality: 135455 bases at least Q30
Consensus quality: 139047 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1125: contig of 1125 bp in length
* 1126 1225: gap of unknown length
* 1226 2448: contig of 1423 bp in length
* 2649 2748: gap of unknown length
* 2749 4281: contig of 1533 bp in length
* 4282 4381: gap of unknown length
* 4382 5618: contig of 1237 bp in length
* 5619 5718: gap of unknown length
* 5719 6752: contig of 1034 bp in length

6753 6852: gap of unknown length
6853 8426: contig of 1574 bp in length
8427 8526: gap of unknown length
8527 9614: contig of 1088 bp in length
9615 9714: gap of unknown length
9715 10916: contig of 1202 bp in length
10917 11016: gap of unknown length
11017 12477: contig of 1461 bp in length
12478 12577: gap of unknown length
12578 14279: contig of 1702 bp in length
14280 14379: gap of unknown length
14380 15623: contig of 1244 bp in length
15624 16975: contig of 1252 bp in length
16976 17075: gap of unknown length
17076 18290: contig of 1215 bp in length
18291 18390: gap of unknown length
18391 19878: contig of 1488 bp in length
19879 19978: gap of unknown length
19979 21805: contig of 1727 bp in length
21806 23139: contig of 1334 bp in length
23140 23239: gap of unknown length
23240 25123: contig of 1884 bp in length
25124 25223: gap of unknown length
25224 26544: contig of 1321 bp in length
26545 26644: gap of unknown length
26645 27787: contig of 1143 bp in length
27788 27887: gap of unknown length
27889 29236: contig of 1349 bp in length
29237 29336: gap of unknown length
29337 30684: contig of 1348 bp in length
30685 30784: gap of unknown length
30785 32375: contig of 1591 bp in length
32376 32475: gap of unknown length
32476 35053: contig of 2578 bp in length
35054 35153: gap of unknown length
35154 37573: contig of 2420 bp in length
37574 37673: gap of unknown length
37674 39912: contig of 2239 bp in length
39913 40012: gap of unknown length
40013 41561: contig of 1549 bp in length
41562 41661: gap of unknown length
41662 43343: contig of 1682 bp in length
43344 43443: gap of unknown length
43444 45205: contig of 1762 bp in length
45206 45305: gap of unknown length
45306 46790: contig of 1485 bp in length
46791 46890: gap of unknown length
46891 48729: contig of 1839 bp in length
48730 48829: gap of unknown length
48830 50597: contig of 1768 bp in length
50598 50697: gap of unknown length
50698 53462: contig of 2765 bp in length
53463 53562: gap of unknown length
53563 55355: contig of 1793 bp in length
55356 55455: gap of unknown length
55456 57987: contig of 2532 bp in length
57988 58087: gap of unknown length
58088 59862: contig of 1775 bp in length
59863 59962: gap of unknown length
59963 62819: contig of 2857 bp in length
62820 62919: gap of unknown length
62920 65169: contig of 2250 bp in length
65170 65269: gap of unknown length
65270 67081: contig of 1812 bp in length
67082 67181: gap of unknown length
67182 69369: contig of 2188 bp in length
69370 69469: gap of unknown length
69470 72381: contig of 2912 bp in length
72382 72481: gap of unknown length
72482 75186: contig of 2705 bp in length
75187 75286: gap of unknown length

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

* 75287 77528: contig of 2242 bp in length
* 77529 77628: gap of unknown length
* 77629 81277: contig of 3649 bp in length
* 81278 81377: gap of unknown length
* 81378 84197: contig of 2820 bp in length
* 84198 84297: gap of unknown length
* 84298 86822: contig of 2525 bp in length
* 86823 89246: gap of unknown length
* 89247 89347: contig of 2324 bp in length
* 89348 91777: gap of unknown length
* 91778 91877: contig of 2431 bp in length
* 91878 93352: contig of 3475 bp in length
* 93353 95452: gap of unknown length
* 95453 99435: contig of 3983 bp in length
* 99436 99535: gap of unknown length
* 99536 106605: contig of 7070 bp in length
* 106606 106705: gap of unknown length
* 106706 111246: contig of 4541 bp in length
* 111247 111346: gap of unknown length
* 111347 114800: contig of 3454 bp in length
* 114801 114900: gap of unknown length
* 114901 119079: contig of 4179 bp in length
* 119080 119179: gap of unknown length
* 119180 123770: contig of 4591 bp in length

Alignment Scores:

Pred. No.:	106	Length:	186357
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	9.62%	Indels:	0
DB:	2	Gaps:	0

US-09-814-661A-2 (1-104) x ACL14454 (1-186357)

Qy 53 SerSerAlaSerAlaSerSerLeu 62

Db 102911 TCCTCCGCTTCGGTTCGGTTCCTCGCTC 102940

RESULT 15

AL671875

LOCUS AL671875 191894 bp DNA linear HTG 19-JUN-2002
Mus musculus chromosome X clone RP23-296M12, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.

ACCESSION AL671875

VERSION AL671875.8 GI:21531190

KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 191894)

REFERENCE Wray, P.

AUTHORS Direct Submission

TITLE Submitted (17-JUN-2002)

Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 21, 2002 this sequence version replaced gi:20502258.

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM296M12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator; 1% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Consensus quality: 191823 bases at least Q40
Consensus quality: 191870 bases at least Q30

Consensus quality: 191881 bases at least Q20
Insert size: 191894; sum-of-contigs
Insert size: 189294; 3.4% error; agarose-fp
Quality coverage: 11.44x in Q20 bases; sum-of-contigs Quality
coverage: 11.76x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

1. 191894

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="X"

/clone="RP23-296M12"

/clone_lib="RPCI-23"

1. 191894

/note="assembly fragment:01249"

misc_feature

BASE COUNT 54652 a 36826 c 36890 g 63566 t

ORIGIN

Alignment Scores:

Pred. No.:	109	Length:	191894
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	9.62%	Indels:	0
DB:	2	Gaps:	0

US-09-814-661A-2 (1-104) x AL671875 (1-191894)

Qy 53 SerSerAlaSerAlaSerAlaSerSerLeu 62

Db 89202 TCCTCTGCATCAGCTTCTGCCTCCAGCTTA 89231

Search completed: July 17, 2003, 20:24:39

Job time : 1488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 18:05:29 ; Search time 64 Seconds
(without alignments)
498.350 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 537

Sequence: 1 MNSQDYFYAQRCCQQOAP.....FGSGELKSMFNGKVEMDF 104

Scoring table: BLOSUMP62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-MODEL=framet_p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09814661/runat.16072003.115310.14331/app_query.fasta_1.263
-DB-Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn.1.1.61/runat.16072003.115310.14331 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	72	13.4	8931	3	US-09-028-934-28
c 2	70.5	13.1	611	4	US-07-757-022B-27
c 3	70.5	13.1	5381	4	US-09-750-580-4
c 4	70.5	13.1	81001	4	US-09-750-580-1
c 5	70.5	13.1	4403765	4	US-09-103-840A-2
c 6	70.5	13.1	4411529	4	US-09-103-840A-1
c 7	70	13.0	111282	4	US-09-754-250-3
c 8	68	12.7	8532	1	US-08-452-655B-1
c 9	68	12.7	8532	3	US-08-450-582-1
c 10	68	12.7	9606	1	US-07-741-940-1
c 11	68	12.7	9606	1	US-08-289-548A-1
c 12	68	12.7	9606	1	US-08-452-654-1

13	68	12.7	9606	2	US-08-370-235A-1	Sequence 1, Appli
14	68	12.7	9606	4	US-08-449-731-1	Sequence 1, Appli
15	67.5	12.6	627	4	US-07-757-022B-93	Sequence 93, Appl
16	67.5	12.6	660	4	US-07-757-022B-95	Sequence 95, Appl
17	67.5	12.6	1584	5	PCT-US92-02091-5	Sequence 5, Appli
18	67.5	12.6	3420	4	US-07-757-022B-103	Sequence 103, App
19	67.5	12.6	4092	4	US-07-757-022B-51	Sequence 51, Appl
20	67.5	12.6	4215	4	US-07-757-022B-61	Sequence 61, Appl
21	67.5	12.6	5008	4	US-07-757-022B-1	Sequence 1, Appli
c 22	66.5	12.4	1066	4	US-09-325-932A-97	Sequence 97, Appl
c 23	65.5	12.2	1337	3	US-08-467-023-1	Sequence 1, Appli
c 24	65	12.1	1734	6	5352575-8	Patent No. 5352575
c 25	65	12.1	2581	4	US-09-370-838-66	Sequence 66, Appl
c 26	65	12.1	5475	2	US-08-680-327-1	Sequence 1, Appli
c 27	65	12.1	5475	4	US-09-228-246-3	Sequence 3, Appli
c 28	65	12.1	10968	2	US-08-680-327-2	Sequence 2, Appli
c 29	65	12.1	10968	4	US-09-228-246-1	Sequence 1, Appli
c 30	64.5	12.0	3252	2	US-08-809-740A-1	Sequence 1, Appli
c 31	64.5	12.0	3252	2	US-08-809-740A-4	Sequence 4, Appli
c 32	63.5	11.8	279	4	US-09-172-108-22	Sequence 22, Appl
c 33	63.5	11.8	279	4	US-09-172-711-20	Sequence 20, Appl
c 34	63.5	11.8	1635	1	US-08-133-347-1	Sequence 1, Appli
c 35	63.5	11.8	1635	1	US-08-133-347-3	Sequence 3, Appli
c 36	63.5	11.8	3251	4	US-09-085-199B-6	Sequence 6, Appli
c 37	63.5	11.8	4377	3	US-08-911-853-28	Sequence 28, Appl
c 38	63.5	11.8	4377	4	US-09-479-409-28	Sequence 28, Appl
c 39	63.5	11.8	4377	4	US-09-479-453-28	Sequence 3, Appli
c 40	63.5	11.8	19736	4	US-09-740-035-3	Sequence 56, Appl
c 41	63	11.7	382	4	US-08-867-611-56	Sequence 51, Appl
c 42	63	11.7	1414	4	US-08-867-611-57	Sequence 57, Appl
c 43	63	11.7	1420	4	US-08-867-611-57	Sequence 1, Appli
c 44	63	11.7	3090	4	US-08-945-983-1	Sequence 131, App
c 45	62.5	11.6	624	4	US-07-757-022B-131	

ALIGNMENTS

RESULT 1
US-09-028-934-28/c
; Sequence 28, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261

```

; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meligs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1506/CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Burkholderia cepacia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 657..2267
; OTHER INFORMATION: /product= "PrnA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2270..3355
; OTHER INFORMATION: /product= "PrnB"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3421..5121
; OTHER INFORMATION: /product= "PrnC"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5145..6266
; OTHER INFORMATION: /product= "PrnD"
;
; US-09-028-934-28
;
; Alignment Scores:
; Pred. No.: 61.8 Length: 8931
; Score: 72.00 Matches: 21
; Percent Similarity: 46.15% Conservative: 21
; Best Local Similarity: 23.08% Mismatches: 33
; Query Match: 13.41% Indels: 16
; DB: 3 Gaps: 3
;
; US-09-814-661A-2 (1-104) x US-09-028-934-28 (1-8931)
;
; Qy 10 AlaGlnAsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThrMetAla 29
; Db 8474 GCCCGGTGCATGTCGCCGGCACATCGCGCCCATCGGGGCAAAA-----8430
;
; Qy 30 GluPheArgGValProLeuProMetAlaGluValProMetLeuSerThrGlnAsn 49
; Db 8429 -----CTCCACCGCGCGCAAAACGTCCTCCCTG-----AAA 8400
;
; Qy 50 SerMetGlySerSerAlaSerAlaSerAlaSerLeuGluMetTrpGluLysAsp---68
; Db 8399 CAGTCGTAAGTGGCGCGCGGCGGCGTGCATCATCGCGCGCGCGCAAAAATCCG 8340
;
; Qy 69 LeuGluGluArgLeuAsnSerIleAspHisAspMetAsnAsnAsnLysPheGlySerGly 88
; Db 8339 ATTCAGGAGACAAAGCAATGCAACGCATCAGGATCAGCCATTCGCGATCGGCACGCGGC 8280
;
; Qy 89 GluLeuLysSerMetPheAsnGlnGlyLysVal 99
; Db 8279 CGCATCGCGCACGCGCTCACGCTGGGACTGTT 8247
;
; RESULT 2
; US-07-757-022B-27
; Sequence 27, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
;
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseiff, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 611 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..333
;
; US-07-757-022B-27
;
; Alignment Scores:
; Pred. No.: 1.82 Length: 611
; Score: 70.50 Matches: 21
; Percent Similarity: 45.16% Conservative: 21
; Best Local Similarity: 22.58% Mismatches: 26
; Query Match: 13.13% Indels: 25
; DB: 4 Gaps: 3
;
; US-09-814-661A-2 (1-104) x US-07-757-022B-27 (1-611)
;
; Qy 17 GlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPhe-----ArgArgVal 34
; Db 291 AAAGAAAGACCTCCACCTTCAGGAGCATCTCAAAACCATCAATAACAAACCAACGTTCA 350
;
; Qy 35 ProLeuProPro-----MetAlaGlu 41
; Db 351 CCCAAACCAACCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 410
;
; Qy 42 ValProMetLeuSerThrGlnAsnSerMetGlySerAlaSerAlaSerAlaSerSer 61
; Db 411 GAACATTCTGTTTCTCTGAAATCAAGAGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 470
```

```
QY 62 LeuGluMetTrpGlu-----LysAspLeuGluGlu 71
      :::::
Db 471 TCAACAATTGGAAATCAAGTCTTCCAAAATTCAGCTGCCTAATAGAGAAATACAGAAG 530
      :::::
QY 72 ArgLeuAsnSerIleAspHisAspMetAsnAsnLys 84
      :::::
Db 531 AAACCTCAAGTAAAGATAACAGAGAAGAACAGAACTAAA 569
      :::::

RESULT 3
US-09-750-580-4
; Sequence 4, Application US/09750580
; Patent No. 6455280
; GENERAL INFORMATION:
; APPLICANT: Yen, Frances
; APPLICANT: Denison, Blake
; APPLICANT: Bour, Barbara
; APPLICANT: Bihain, Bernard
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Ebbets-Reed, Dana
; APPLICANT: Salter-Cld, Luisa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
; FILE REFERENCE: 89.US2.CIP
; CURRENT APPLICATION NUMBER: US/09/750,580
; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 09/599,362
; PRIOR FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: PCT/IB00/0101
; PRIOR FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: PCT/IB99/02058
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 49/469/099
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: US 60/113,686
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 60/141,032
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 5381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..918
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 919..930
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 1442..1498
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 1613..1724
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 2243..3940
; OTHER INFORMATION: exon 4
; NAME/KEY: misc_feature
; LOCATION: 3941..5381
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 319
; OTHER INFORMATION: 17-42-319 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 3213
; OTHER INFORMATION: 17-41-250 : polymorphic base C or T
; NAME/KEY: conflict
; LOCATION: 1241
; OTHER INFORMATION: 17-39-343 : T in ref genbank AC007707

; NAME/KEY: conflict
; LOCATION: 1447
; OTHER INFORMATION: 17-40-202 : G in ref genbank AC007707
; NAME/KEY: primer_bind
; LOCATION: 1..11022
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 553..11575
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 899..11920
; OTHER INFORMATION: 17-39.pu
; NAME/KEY: primer_bind
; LOCATION: 1246..12267
; OTHER INFORMATION: 17-40.pu
; NAME/KEY: primer_bind
; LOCATION: 1441..12461
; OTHER INFORMATION: 17-39.rp complement
; NAME/KEY: primer_bind
; LOCATION: 1632..12651
; OTHER INFORMATION: 17-40.rp complement
; NAME/KEY: primer_bind
; LOCATION: 2964..13984
; OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 3432..14454
; OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 300..318
; OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer_bind
; LOCATION: 320..338
; OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer_bind
; LOCATION: 3194..3212
; OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer_bind
; LOCATION: 3214..3232
; OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: misc_binding
; LOCATION: 307..331
; OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc_binding
; LOCATION: 3201..3225
; OTHER INFORMATION: 17-41-250.probe
; US-09-750-580-4

Alignment Scores:
Pred. No.: 46.2 Length: 5381
Score: 70.50 Matches: 23
Percent Similarity: 45.95% Conservative: 11
Best Local Similarity: 31.08% Mismatches: 21
Query Match: 13.13% Indels: 19
DB: 4 Gaps: 4

US-09-814-661a-2 (1-104) x US-09-750-580-4 (1-5381)
QY 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro 37
      ||| |||||
Db 2109 CAAGAGCCCTCGACCTGGGG-----CCAAGCC 2138
      ||| |||

QY 38 ProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAla 57
      :::::
Db 2139 TGGCTCTTGGTTCCCGGAGGATCAGTCGGCGACTGTGGGACAAAGGAGATGATC 2198
      :::::
QY 58 SerAlaSerSerLeuGluMetTrp-----GluLysAspLeuGluArg 72
      |||||
Db 2199 GGGCTAGCAGTCTGACGGCTGTGATATCTGCCCTTCTCCAGGACCTGAAAGAC--- 2255
      |||||
QY 73 LeuAsnSerIleAspHisAspMetAsnAsn-----AsnLysPhe 85
      |||||
Db 2256 -----AGCCTTGAGCAAGACCTCAACAATATGAACAAGTTC 2291
      |||||
```

```

RESULT 4
US-09-750-580-1
: Sequence 1, Application US/09750580
: Patent No. 6455280
: GENERAL INFORMATION:
: APPLICANT: Yen, Frances
: APPLICANT: Denison, Blake
: APPLICANT: Bour, Barbara
: APPLICANT: Bihain, Bernard
: APPLICANT: Dumas Milne Edwards, Jean-Baptiste
: APPLICANT: Duclert, Aymeric
: APPLICANT: Bouqueleret, Lydie
: APPLICANT: Ebbets-Reed, Dana
: APPLICANT: Salter-Cid, Luisa
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
: FILE REFERENCE: 89 US2.CIP
: CURRENT APPLICATION NUMBER: US/09/750,580
: PRIOR FILING DATE: 2000-12-28
: PRIOR APPLICATION NUMBER: US 09/599,362
: PRIOR FILING DATE: 2000-06-21
: PRIOR APPLICATION NUMBER: PCT/IB00/0101
: PRIOR FILING DATE: 2000-06-21
: PRIOR APPLICATION NUMBER: PCT/IB99/02058
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: US 49/469/099
: PRIOR FILING DATE: 1999-12-21
: PRIOR APPLICATION NUMBER: US 60/113,686
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: US 60/141,032
: PRIOR FILING DATE: 1999-06-25
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Patent.pm
: SEQ ID NO 1
: LENGTH: 81001
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 10946..12946
: OTHER INFORMATION: 5'regulatory region
: NAME/KEY: exon
: LOCATION: 12947..12958
: OTHER INFORMATION: exon 1
: NAME/KEY: exon
: LOCATION: 13470..13526
: OTHER INFORMATION: exon 2
: NAME/KEY: exon
: LOCATION: 13641..13752
: OTHER INFORMATION: exon 3
: NAME/KEY: exon
: LOCATION: 14271..15968
: OTHER INFORMATION: exon 4
: NAME/KEY: misc_feature
: LOCATION: 15969..17969
: OTHER INFORMATION: 3'regulatory region
: NAME/KEY: allele
: LOCATION: 1239
: OTHER INFORMATION: 20-828-311 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 12347
: OTHER INFORMATION: 17-42-319 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 15241
: OTHER INFORMATION: 17-41-250 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 42218
: OTHER INFORMATION: 20-841-149 : polymorphic base A or G
: NAME/KEY: allele
: LOCATION: 45442
: OTHER INFORMATION: 20-842-115 : polymorphic base A or G
: NAME/KEY: allele
: LOCATION: 77058
: OTHER INFORMATION: 20-853-415 : polymorphic base C or T

```

```

: NAME/KEY: primer_bind
: LOCATION: 929..949
: OTHER INFORMATION: 20-828.pu
: NAME/KEY: primer_bind
: LOCATION: 1357..1377
: OTHER INFORMATION: 20-828.rp complement
: NAME/KEY: primer_bind
: LOCATION: 12029..12050
: OTHER INFORMATION: 17-42.pu
: NAME/KEY: primer_bind
: LOCATION: 12581..12603
: OTHER INFORMATION: 17-42.rp complement
: NAME/KEY: primer_bind
: LOCATION: 14992..15012
: OTHER INFORMATION: 17-41.pu
: NAME/KEY: primer_bind
: LOCATION: 15460..15482
: OTHER INFORMATION: 17-41.rp complement
: NAME/KEY: primer_bind
: LOCATION: 42070..42090
: OTHER INFORMATION: 20-841.pu
: NAME/KEY: primer_bind
: LOCATION: 42572..42591
: OTHER INFORMATION: 20-841.rp complement
: NAME/KEY: primer_bind
: LOCATION: 45328..45347
: OTHER INFORMATION: 20-842.pu
: NAME/KEY: primer_bind
: LOCATION: 45863..45883
: OTHER INFORMATION: 20-842.rp complement
: NAME/KEY: primer_bind
: LOCATION: 76644..76664
: OTHER INFORMATION: 20-853.pu
: NAME/KEY: primer_bind
: LOCATION: 77166..77185
: OTHER INFORMATION: 20-853.rp complement
: NAME/KEY: primer_bind
: LOCATION: 1220..1238
: OTHER INFORMATION: 20-828-311.mis
: NAME/KEY: primer_bind
: LOCATION: 1240..1258
: OTHER INFORMATION: 20-828-311.mis complement
: NAME/KEY: primer_bind
: LOCATION: 12328..12346
: OTHER INFORMATION: 17-42-319.mis
: NAME/KEY: primer_bind
: LOCATION: 12348..12366
: OTHER INFORMATION: 17-42-319.mis complement
: NAME/KEY: primer_bind
: LOCATION: 15222..15240
: OTHER INFORMATION: 17-41-250.mis
: NAME/KEY: primer_bind
: LOCATION: 15242..15260
: OTHER INFORMATION: 17-41-250.mis complement
: NAME/KEY: primer_bind
: LOCATION: 42199..42217
: OTHER INFORMATION: 20-841-149.mis
: NAME/KEY: primer_bind
: LOCATION: 42219..42237
: OTHER INFORMATION: 20-841-149.mis complement
: NAME/KEY: primer_bind
: LOCATION: 45423..45441
: OTHER INFORMATION: 20-842-115.mis
: NAME/KEY: primer_bind
: LOCATION: 45443..45461
: OTHER INFORMATION: 20-842-115.mis complement
: NAME/KEY: primer_bind
: LOCATION: 77039..77057
: OTHER INFORMATION: 20-853-415.mis
: NAME/KEY: primer_bind
: LOCATION: 77059..77077
: OTHER INFORMATION: 20-853-415.mis complement
: NAME/KEY: misc_binding

```



```
; LOCATION: 1227..1251
; OTHER INFORMATION: 20-828-311.probe
; NAME/KEY: misc_binding
; LOCATION: 12335..12359
; OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc_binding
; LOCATION: 15229..15253
; OTHER INFORMATION: 17-41-250.probe
; NAME/KEY: misc_binding
; LOCATION: 42206..42230
; OTHER INFORMATION: 20-841-149.probe
; NAME/KEY: misc_binding
; LOCATION: 45430..45454
; OTHER INFORMATION: 20-842-115.probe
; NAME/KEY: misc_binding
; LOCATION: 77046..77070
; OTHER INFORMATION: 20-853-415.probe
; US-09-750-580-1
Alignment Scores:
Pred. No.: 2.6e+03      Length: 81001
Score: 70.50           Matches: 23
Percent Similarity: 45.95%      Conservative: 11
Best Local Similarity: 31.08%    Mismatches: 21
Query Match: 13.13%             Indels: 19
DB: 4                     Gaps: 4

US-09-814-661A-2 (1-104) x US-09-750-580-1 (1-81001)
Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro 37
Db 14137 CAAGAGCCCTCGACCTCGGG-----CCAAAGCC 14166
Qy 38 ProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAla 57
Db 14167 TGGCTCTTGTCCCCAGAGGATCAGTCGCGGATGACTGGGGACAAAGGAGATGATG 14226
Qy 58 SerAlaSerSerLeuGluMetTrp-----GluLysAspLeuGluGluArg 72
Db 14227 GGGCTAGCAGTCTGACGGCTGTGATATCTGCCCTTCTCCAGGACCTGAAGAC--- 14283
Qy 73 LeuAsnSerLeuAspHisAspMetAsnAsn-----AsnLysPhe 85
Db 14284 -----AGCCTTGAGCAAGACCTCAACAATATGAACAAGTTC 14319

RESULT 5
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 5.93e+05      Length: 4411529
Score: 70.50           Matches: 27
Percent Similarity: 40.68%      Conservative: 21
Best Local Similarity: 22.88%    Mismatches: 42
Query Match: 13.13%             Indels: 28
DB: 4                     Gaps: 3

US-09-814-661A-2 (1-104) x US-09-103-840A-1 (1-4411529)
Qy 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGln-----GlnGlnGln 18
Db 3034389 CTTCCGCGCGGCAAGATGTTTTTCACCCCGCCGCGGTCGCAATACCTCTACGCGCC 3034330
Qy 19 AlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPr 38
Db 3034329 GCTCCTGGACAAATACGAGTCTGCGGCGGTGCGAGCAGCATCTGGGATTC 3034270

us-09-814-661a-2 (1-104) x US-09-103-840A-2 (1-4403765)
Qy 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGln-----GlnGlnGln 18
Db 3029288 CTTCCGCGCGGCAAGATGTTTTTCACCCCGCCGCGGTCGCAATACCTCTACGCGCC 3029229
Qy 19 AlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPr 38
Db 3029228 GCTCCTGGACAAATACGAGTCTGCGGCGGTGCGAGCAGCATCTGGGATTCGCGAC 3029169
Qy 38 oMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSe 58
Db 3029168 CATCAACGAGGTCCCGGCTTGCAGGACCGCAGTCTGCGGTGAGCAGTCTCCGCGGC 3029109
Qy 58 rAlaSerSerLeuGluMetTrpGluLysAsp-----LeuGluGluArgLe 73
Db 3029108 CGACACCCCGGTGATTTTCTGGACGCTGAACACACAGGTTCGTCGTGCGGCGCGCT 3029049
Qy 73 uAsn----- 74
Db 3029048 GAATGCCGCTGGGACAAAGCTCGGCAGCTCGGTGGCGTCTGGTGCCTCGGTCGCGG 3028989
Qy 75 ----SerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeu 90
Db 3028988 TGAACCTACGAGCGGAGGTCAACGCCAGAGTTTCAGCGGTGTTGAAGTC 3028937

RESULT 6
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 5.93e+05      Length: 4411529
Score: 70.50           Matches: 27
Percent Similarity: 40.68%      Conservative: 21
Best Local Similarity: 22.88%    Mismatches: 42
Query Match: 13.13%             Indels: 28
DB: 4                     Gaps: 3

US-09-814-661A-2 (1-104) x US-09-103-840A-1 (1-4411529)
Qy 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGln-----GlnGlnGln 18
Db 3034389 CTTCCGCGCGGCAAGATGTTTTTCACCCCGCCGCGGTCGCAATACCTCTACGCGCC 3034330
Qy 19 AlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPr 38
Db 3034329 GCTCCTGGACAAATACGAGTCTGCGGCGGTGCGAGCAGCATCTGGGATTC 3034270
```

```

QY 38 oMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAla 58
Db 3034269 CATCAACGAGGTCGCCGCGCTTCCGACCGACAGCTCGGTGAGCACGTTCTCCGCGC 3034210
QY 58 rAlaSerSerLeuGluMetTrpGluLysAsp-----LeuGluGluArgLe 73
Db 3034209 CGAACACCGCGTGATTTCTGGACGCTGAACACACGCGTGGTGGTGGCGCGCT 3034150
QY 73 uAsp----- 74
Db 3034149 GAATGCCGCGTGGGACAGCTCGGACGCTCGGCGTGGTGGTGGTGGTGGTGGGGA 3034090
QY 75 -----SerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeu 90
Db 3034089 TGAACCTACGACGCGGAGTCAACCGCCAGAGTTACGCGGTGGTGAAGTC 3034038

```

RESULT 7

```

US-09-754-250-3/c
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL001063
; CURRENT APPLICATION NUMBER: US/09/754,250
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(111282)
; OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3

```

```

Alignment Scores:
Pred. No.: 4.87e+03 Length: 111282
Score: 70.00 Matches: 24
Percent Similarity: 42.50% Conservative: 10
Best Local Similarity: 30.00% Mismatches: 25
Query Match: 13.04% Indels: 21
DB: 4 Gaps: 4

```

US-09-814-661A-2 (1-104) x US-09-754-250-3 (1-111282)

```

QY 14 CysGlnGlnGlnAlaProSerThr-----LeuArgThrValThrMetAla 29
Db 489 TGCCAGGCACACAGCCCTCCCTTAACCAATAGGAGGGGACTGAAGTGCAGAGGTGAG 430
QY 30 GluPheArg-----ArgValProLeuProMetAla----- 40
Db 429 GAGATTCGGGACGCTGGGTGCCCTTTCACTGTGCTCCCTCACTCCATCAGCCCTATGAGGT 370
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSer 60
Db 369 CAGGTCCTCCATCCTGAGCTCTCAGGAGAGCGAGGAGGAGGAGTTCCTCAGTCA 310
QY 61 SerLeu-Glu-----MetTrpGluLysAspLeuGluGluArg 72
Db 309 AGAGTAAACAGAGAGGCGCTCAGAGTGACTTGTGTGGGCGGAGATACACAGGAGCC 252

```

RESULT 8

```

US-08-452-655B-1
; Sequence 1, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH

```

```

; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP2.5 (APC)
; US-08-452-655B-1
Alignment Scores:
Pred. No.: 198 Length: 8532
Score: 68.00 Matches: 27
Percent Similarity: 45.36% Conservative: 17
Best Local Similarity: 27.84% Mismatches: 33
Query Match: 12.66% Indels: 20
DB: 1 Gaps: 3
US-09-814-661A-2 (1-104) x US-08-452-655B-1 (1-8532)
QY 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro 37
Db 7600 GAAAGTCCTTCTAGACTTCCAATAGGTGAGCAACCTGGAAAGCTGAGCACACAAA 7659
QY 38 ProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerAla----- 55
Db 7660 CATTTCATCCCTTCTCGAGTACCACTTCGAGAGAGAACTGGAACTTCATCTTCATT 7719
QY 56 ---SerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74

```

Db 7720 CTTTCGTTTCATCAGAAATCCAGTGAAAAAGCAAAAGCAAAAGTGAAGATGAAACATGTGAAC 7779
Qy 75 SerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLysSerMetPhe 94
Db 7780 TCTATT-----TCAGGAACCAACAAGTAAGAA 7809
Qy 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
Db 7810 AACCAAGTATCCGAAAAGGAACATGGAGAAAAATAAAGAAATGAATTT 7860
RESULT 9
US-08-450-582-1
; Sequence 1, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP2.5(APC)
US-08-450-582-1

Alignment Scores:

Pred. No.: 198 Length: 8532
Score: 68.00 Matches: 27
Percent Similarity: 45.36% Conservative: 17
Best Local Similarity: 27.84% Mismatches: 33
Query Match: 12.66% Indels: 20
DB: 3 Gaps: 3
US-09-814-661A-2 (1-104) x US-08-450-582-1 (1-8532)
Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro 37
Db 7600 GAAAGTCCTCTAGACTTCCTCAATCAATAGTCAGGAACCTGAAACGGTGAGCACAGCAA 7659
Qy 38 ProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAla 55
Db 7660 CATTCAATCATCCCTTCCTCGAGTAGCACTTGGAGAGAAGACTGGAAGTTTCATCTTCAATT 7719
Qy 56 ---SerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74
Db 7720 CTTTCTGCTTCATCAGAAATCCAGTGAAAAAGCAAAAGTGAAGATGAAACATGTGAAC 7779
Qy 75 SerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLysSerMetPhe 94
Db 7780 TCTATT-----TCAGGAACCAACAAGTAAGAA 7809
Qy 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
Db 7810 AACCAAGTATCCGAAAAGGAACATGGAGAAAAATAAAGAAATGAATTT 7860
RESULT 10
US-07-741-940-1
; Sequence 1, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9606 base pairs

```
;
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP2.5(APC)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..8562
US-07-741-940-1
Alignment Scores:
Pred. No.: 236 Length: 9606
Score: 68.00 Matches: 27
Percent Similarity: 45.36% Conservative: 17
Best Local Similarity: 27.84% Mismatches: 33
Query Match: 12.66% Indels: 20
DB: 1 Gaps: 3
US-09-814-661A-2 (1-104) x US-07-741-940-1 (1-9606)
Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuPro 37
Db 7633 GAAAGTCCTCTAGACTTCCAAATCAATAGGTGAGAACCTGGAAACGTGAGCAGCAAAA 7692
Qy 38 ProMetAlaGluValProMetLeuSerThrClnAsnSerMetGlySerSerAla----- 55
Db 7693 CATTCATCATCCCTTCCTCGAGTAAGCACCTGGAGAACACTGGAAGTTCATCTCAATT 7752
Qy 56 ---SerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74
Db 7753 CTTTCTGCTTCATCAGATCCAGTGAAACCAAGAGTGGAGATGAGGATGAAACATGTGAC 7812
Qy 75 SerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLeuLysSerMetPhe 94
Db 7813 TCTATT-----TCAGGAACCAACAAAGTAAAGAA 7842
Qy 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
Db 7843 AACCAAGTATCCGCAAAAGGAACATGGAGAAAATAAAGAAATGAATTT 7893
RESULT 11
US-08-289-548A-1
; Sequence 1, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allogretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9606 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: DP2.5(APC)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..8562
US-08-289-548A-1
Alignment Scores:
Pred. No.: 236 Length: 9606
Score: 68.00 Matches: 27
Percent Similarity: 45.36% Conservative: 17
Best Local Similarity: 27.84% Mismatches: 33
Query Match: 12.66% Indels: 20
DB: 1 Gaps: 3
US-09-814-661A-2 (1-104) x US-08-289-548A-1 (1-9606)
Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuPro 37
Db 7633 GAAAGTCCTCTAGACTTCCAAATCAATAGGTGAGAACCTGGAAACGTGAGCAGCAAAA 7692
Qy 38 ProMetAlaGluValProMetLeuSerThrClnAsnSerMetGlySerSerAla----- 55
Db 7693 CATTCATCATCCCTTCCTCGAGTAAGCACCTGGAGAACACTGGAAGTTCATCTCAATT 7752
Qy 56 ---SerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74
Db 7753 CTTTCTGCTTCATCAGATCCAGTGAAACCAAGAGTGGAGATGAGGATGAAACATGTGAC 7812
Qy 75 SerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLeuLysSerMetPhe 94
Db 7813 TCTATT-----TCAGGAACCAACAAAGTAAAGAA 7842
Qy 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
Db 7843 AACCAAGTATCCGCAAAAGGAACATGGAGAAAATAAAGAAATGAATTT 7893
RESULT 12
US-08-452-654-1
; Sequence 1, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
```

```

: TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner, Birch, McKie & Beckett
: STREET: 1001 G Street, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20001-4598
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/452,654
: FILING DATE: 25-MAY-1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/741,940
: FILING DATE: 08-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 1107.035574
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9606 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: DP2.5(APC)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 34..8562
: US-08-452-654-1
:
: Alignment Scores:
: Pred. No.: 236 Length: 9606
: Score: 68.00 Matches: 27
: Percent Similarity: 45.36% Conservative: 17
: Best Local Similarity: 27.84% Mismatches: 33
: Query Match: 12.66% Indels: 20
: DB: 1 Gaps: 3
:
: US-09-814-661A-2 (1-104) x US-08-452-654-1 (1-9606)
:
: Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuPro 37
: Db 7633 GAAAGTCCTTCTAGACTTCCAATCAATAGGTTCAGGAACCTGGAAACCTGAGCAGCAAA 7692
:
: Qy 38 ProMetAlaGluValProMetLeuSerThrClnAsnSerMetGlySerSerAla----- 55
: Db 7693 CATTTCATCCCTTCCTCGAGTAGGACCTGGAGAGAAGCTGGGAAGCTTCATCTCAATT 7752
:
: Qy 56 ---SerAlaSerAlaSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74
: Db 7753 CTTTCTGCTTCATCAGATCCAGTGAAAGCAAAAGTGGAGATGAAACATGTGAAC 7812
:
: Qy 75 SerIleAspHisMetAsnAsnLysPheGlySerGlyGluLeuLysSerMetPhe 94
: Db 7813 TCTATT-----TCAGGACCAACCAAGTAAAGAA 7842
:
: Qy 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
: Db 7843 AACCAAGTATCCGCAAAAGGACATGGAGAAATAAAGAAATGAATTT 7893

```

```

RESULT 13
US-08-370-235A-1
: Sequence 1, Application US/08370235A
: Patent No. 5910418
: GENERAL INFORMATION:
: APPLICANT: VOGELSTEIN, BERT
: APPLICANT: KINZLER, KENNETH W.
: APPLICANT: HILL, DAVID E.
: APPLICANT: JOHNSON, KAREN A.
: TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING
: TITLE OF INVENTION: MUTATIONS IN THE APC GENE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BANNER & WITCOFF, LTD.
: STREET: 1001 G STREET, N.W.
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20001
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/370.235A
: FILING DATE: 01-JAN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: KAGAN, SARAH A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 01107.48688
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 508 9100
: TELEFAX: 202 508 9299
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9606 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: POSITION IN GENOME:
: MAP POSITION: 5q21
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 34..8562
: US-08-370-235A-1
:
: Alignment Scores:
: Pred. No.: 236 Length: 9606
: Score: 68.00 Matches: 27
: Percent Similarity: 45.36% Conservative: 17
: Best Local Similarity: 27.84% Mismatches: 33
: Query Match: 12.66% Indels: 20
: DB: 2 Gaps: 3
:
: US-09-814-661A-2 (1-104) x US-08-370-235A-1 (1-9606)
:
: Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuPro 37
: Db 7633 GAAAGTCCTTCTAGACTTCCAATCAATAGGTTCAGGAACCTGGAAACCTGAGCAGCAAA 7692
:
: Qy 38 ProMetAlaGluValProMetLeuSerThrClnAsnSerMetGlySerSerAla----- 55
: Db 7693 CATTTCATCCCTTCCTCGAGTAGGACCTGGAGAGAAGCTGGGAAGCTTCATCTCAATT 7752
:
: Qy 56 ---SerAlaSerAlaSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74

```

Db 7753 CTTTCTGCTTCATCAGATCCAGTGAAGCAAAAGTGAAGATGAGGATGGAAC 7812
QY 75 SerileAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLysSerMetPhe 94
Db 7813 TCTATT-----TCAGGAACCAACAAAGTAAAGAA 7842
QY 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
Db 7843 AACCAAGTATCCGCAAAAGACATGGAGAAAATAAAGAAATGAATTT 7893
RESULT 14
US-08-449-731-1
Sequence 1, Application US/08449731
Patent No. 6413727
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
ANAND, RAKESH
CARLSON, MARY
GRODEN, JOANNA
HEDGE, PHILIP J.
JOSLYN, GEOFF
KINZLER, KENNETH
MARKHAM, ALEXANDER F.
NAKAMURA, YUSUKE
THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESS: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9606 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CLONE: DP2.5(APC)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..8562
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-449-731-1
Alignment Scores:

Pred. No.: 236 Length: 9606
Score: 68.00 Matches: 27
Percent Similarity: 45.36% Conservative: 17
Best Local Similarity: 27.84% Mismatches: 33
Query Match: 12.66% Indels: 20
DB: 4 Gaps: 3
US-09-814-661A-2 (1-104) x US-08-449-731-1 (1-9606)
QY 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuPro 37
Db 7633 GAAAGTCTCTAGACTTCCAAATAGGTGAGGAACCTGCAAAACGTGAGCACAGCAAA 7692
QY 36 ProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlyLysSerAla----- 55
Db 7693 CATTCATCATCCCTTCCTCGAGTAAGCACCTGCGAAGAACTGGAAGTTCATCTTCAATT 7752
QY 56 ---SerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsn 74
Db 7753 CTTTCTGCTTCATCAGATCCAGTCAAAAAGCAAAAGTGAAGATGAAAACATGTGAAC 7812
QY 75 SerileAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLysSerMetPhe 94
Db 7813 TCTATT-----TCAGGAACCAACAAAGTAAAGAA 7842
QY 95 AsnGln-----GlyLysValGluGluMetAspPhe 104
Db 7843 AACCAAGTATCCGCAAAAGACATGGAGAAAATAAAGAAATGAATTT 7893
RESULT 15
US-07-757-022B-93
Sequence 93, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:

Search completed: July 17, 2003, 19:18:15
Job time : 1076 secs

THIS PAGE BLANK (OPTIONAL)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:41:21 ; Search time 26 seconds
(without alignments)
117.692 Million cell updates/sec

Title: US-09-814-661a-2

Perfect score: 537

Sequence: 1 MONSDYFYAQRNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	12.7	2842	1	US-07-741-940-7
2	68	12.7	2842	1	US-08-289-548A-7
3	68	12.7	2842	1	US-08-452-554-7
4	68	12.7	2842	4	US-08-449-731-7
5	68	12.7	2843	1	US-07-741-940-7
6	68	12.7	2843	1	US-08-289-548A-2
7	68	12.7	2843	1	US-08-452-554-2
8	68	12.7	2843	1	US-08-452-555B-2
9	68	12.7	2843	1	US-08-452-555B-2
10	68	12.7	2843	1	US-08-370-235A-2
11	68	12.7	2843	3	US-08-450-582-2
12	68	12.7	2843	3	US-08-450-582-7
13	68	12.7	2843	4	US-08-449-731-2
14	68	12.7	2973	2	US-08-821-355A-7
15	68	12.7	2973	2	US-09-003-687A-7
16	68	12.7	2973	4	US-09-136-605-7
17	67.5	12.6	209	4	US-07-757-022B-94
18	67.5	12.6	220	4	US-07-757-022B-96
19	67.5	12.6	1140	4	US-07-757-022B-104
20	67.5	12.6	1363	4	US-07-757-022B-52
21	67.5	12.6	1404	4	US-07-757-022B-2
22	67.5	12.6	1404	4	US-07-757-022B-62
23	65	12.1	577	6	5352575-9
24	65	12.1	1824	2	US-08-680-327-3
25	65	12.1	1824	4	US-09-228-246-2
26	63	11.7	806	4	US-08-945-983-2
27	62.5	11.6	208	4	US-07-757-022B-132

28 62.5 11.6 334 6 5290690-10
29 62.5 11.6 335 6 5290690-9
30 62.5 11.6 587 4 US-08-481-190-4
31 62.5 11.6 587 5 PCT-US93-00869-4
32 61.5 11.5 543 2 US-08-469-412A-7
33 61.5 11.5 543 4 US-09-021-715-7
34 61.5 11.5 734 4 US-09-185-160-13
35 61 11.4 196 2 US-08-829-110-5
36 61 11.4 196 2 US-08-748-483-3
37 61 11.4 230 1 US-08-118-469A-3
38 61 11.4 230 1 US-08-909-119-3
39 61 11.4 380 4 US-09-298-886-8
40 61 11.4 449 3 US-08-987-743-7
41 60.5 11.3 496 2 US-08-696-349-2
42 60.5 11.3 496 5 PCT-US96-13156-2
43 60.5 11.3 555 3 US-08-968-563-34
44 60.5 11.3 555 4 US-08-969-683A-34
45 60 11.2 788 1 US-08-194-338-12

ALIGNMENTS

RESULT 1
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESS: Banner, Birch, Mckie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07741.940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens

Patent No. 5290690
Patent No. 5290690
Sequence 4, Appli
Sequence 4, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 13, Appli
Sequence 5, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 8, Appli
Sequence 7, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 34, Appli
Sequence 34, Appli
Sequence 12, Appli

```
; IMMEDIATE SOURCE:
; CLONE: APC
US-07-741-940-7

Query Match      12.7%; Score 68; DB 1; Length 2842;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSMGSSA---SASASSLEMEKDLLEERLN 74
DB 2533 ESPSRLPINRSGTWKREHSHSSLPVSTWRTGSSSILSASSESEKAKSEDEKHVN 2592
QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
DB 2593 SI-----SGTKQSKENQVSAKGTWRKIKENEK 2619

RESULT 2
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-289-548A-7

Query Match      12.7%; Score 68; DB 1; Length 2842;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSMGSSA---SASASSLEMEKDLLEERLN 74
DB 2533 ESPSRLPINRSGTWKREHSHSSLPVSTWRTGSSSILSASSESEKAKSEDEKHVN 2592
QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
DB 2593 SI-----SGTKQSKENQVSAKGTWRKIKENEK 2619

RESULT 3
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-452-654-7

Query Match      12.7%; Score 68; DB 1; Length 2842;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSMGSSA---SASASSLEMEKDLLEERLN 74
```

Db 2533 ESPSLPINRSGTWKREHSKIUSSSLPRVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2592
QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
Db 2593 SI-----SGTKSKENQVSAKGTWRKIKENE 2619

RESULT 4
US-08-449-731-7
; Sequence 7, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; ANAND, RAKESH
; CARLSON, MARY
; GRODEN, JOANNA
; HEDGE, PHILIP J.
; JOSLYN, GEOFF
; KINZLER, KENNETH
; MARKHAM, ALEXANDER F.
; NAKAMURA, YUSUKE
; THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,731
; FILING DATE: 25-May-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,548
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Query Match 12.7%; Score 68; DB 4; Length 2842;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;
QY 18 QAPSTLRTVTMAEPRRVLPPMAEVPMLSTQNSMGSSA---SASASSLEMMWEKDLERLN 74
Db 2533 ESPSLPINRSGTWKREHSKIUSSSLPRVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2592

QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
Db 2593 SI-----SGTKSKENQVSAKGTWRKIKENE 2619
RESULT 5
US-07-741-940-2
; Sequence 2, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-741-940-2

Query Match 12.7%; Score 68; DB 1; Length 2843;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;
QY 18 QAPSTLRTVTMAEPRRVLPPMAEVPMLSTQNSMGSSA---SASASSLEMMWEKDLERLN 74
Db 2534 ESPSLPINRSGTWKREHSKIUSSSLPRVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2593
QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
Db 2594 SI-----SGTKSKENQVSAKGTWRKIKENE 2620

RESULT 6
US-08-289-548A-2
; Sequence 2, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERTIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-289-548A-2

Query Match 12.7% Score 68; DB 1; Length 2843;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;
Qy 18 QAPSTLRTVTVAEFRVPLPPMAEVPMLSTONSGSSA---SASASSLEWKEKDLERLN 74
Db 2534 ESPRLPINSRGTWKREHSHSSLPVSTWRTTGGSSSILSASSESEKAKSEDEKHVN 2593
Qy 75 SIDHDMNNKFGSGELKSMFNQ-----GKVEEMDF 104
Db 2594 SI-----SGTKSKENQVSAKGTWRKIKENEF 2620

RESULT 7
US-08-452-654-2
Sequence 2, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERTIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-654-2

Query Match 12.7% Score 68; DB 1; Length 2843;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;
Qy 18 QAPSTLRTVTVAEFRVPLPPMAEVPMLSTONSGSSA---SASASSLEWKEKDLERLN 74
Db 2534 ESPRLPINSRGTWKREHSHSSLPVSTWRTTGGSSSILSASSESEKAKSEDEKHVN 2593
Qy 75 SIDHDMNNKFGSGELKSMFNQ-----GKVEEMDF 104
Db 2594 SI-----SGTKSKENQVSAKGTWRKIKENEF 2620

RESULT 8
US-08-452-655B-2
Sequence 2, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERTIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US/08/452.655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
APPLICATION NUMBER: 1107.49964
REFERENCE/DOCKET NUMBER: 32.141
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-655B-2

Query Match 12.7%; Score 68; DB 1; Length 2843;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVPLPPMAEYPMILSTQNSMGSSA---SASASSLEWWEKDLERLN 74
DB 2534 ESPSLRPLNRSGTWKREHSKHSSLPVSTWRTGSSSILSASSESEKAKSEDEKHVN 2593
QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
DB 2594 SI-----SGTKQSKENQVSAKGTWRKIKENE 2620

RESULT 9
US-08-452-655B-7
Sequence 7, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODIN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452.655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
APPLICATION NUMBER: 1107.49964
REFERENCE/DOCKET NUMBER: 32.141
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-452-655B-7

Query Match 12.7%; Score 68; DB 1; Length 2843;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVPLPPMAEYPMILSTQNSMGSSA---SASASSLEWWEKDLERLN 74
DB 2534 ESPSLRPLNRSGTWKREHSKHSSLPVSTWRTGSSSILSASSESEKAKSEDEKHVN 2593
QY 75 SIDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
DB 2594 SI-----SGTKQSKENQVSAKGTWRKIKENE 2620

RESULT 10
US-08-370-235A-2
Sequence 2, Application US/08370235A
Patent No. 5910418
GENERAL INFORMATION:
APPLICANT: VOGELSTEIN, BERT
APPLICANT: KINZLER, KENNETH W.
APPLICANT: HILL, DAVID E.
APPLICANT: JOHNSON, KAREN A.
TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: BANNER & WITCOFF, LTD.
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370.235A
FILING DATE: 01-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32.141

REFERENCE/DOCKET NUMBER: 01107.48688
TELEPHONE: 202 508 9100
TELEFAX: 202 508 9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-370-235A-2

Query Match 12.7%; Score 68; DB 2; Length 2843;

Best Local Similarity 27.8%; Pred. No. 44;

Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRVTMAEPRRVLPPMAEVPMLSTQNSMCSA---SASASSLEMEKDLERLN 74
Db 2534 ESPSLRPNRSCTWKREHSHSSSLPRVSTWRTTGGSSSISSASSESEKAKSEDEKHVN 2593
QY 75 STDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
Db 2594 SI-----SGTKQSKENQVSAKGTWRKIKENEF 2620

RESULT 11

US-08-450-582-2

Sequence 2, Application US/08450582

Patent No. 6114124

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450.582

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/452,655

FILING DATE: 25-MAY-1995

APPLICATION NUMBER: US 08/289,548

FILING DATE: 12-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.49964

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-450-582-2

Query Match 12.7%; Score 68; DB 3; Length 2843;

Best Local Similarity 27.8%; Pred. No. 44;

Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRVTMAEPRRVLPPMAEVPMLSTQNSMCSA---SASASSLEMEKDLERLN 74
Db 2534 ESPSLRPNRSCTWKREHSHSSSLPRVSTWRTTGGSSSISSASSESEKAKSEDEKHVN 2593
QY 75 STDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
Db 2594 SI-----SGTKQSKENQVSAKGTWRKIKENEF 2620

RESULT 12

US-08-450-582-7

Sequence 7, Application US/08450582

Patent No. 6114124

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450.582

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/452,655

FILING DATE: 25-MAY-1995

APPLICATION NUMBER: US 08/289,548

FILING DATE: 12-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.49964

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
US-08-450-582-7

```

```

Query Match      12.7%; Score 68; DB 3; Length 2843;
Best Local Similarity 27.8%; Pred. NO. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVRPLPPMAEYVPLSTQNSGSSA-----SASASLEMMEXKDLERLN 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2534 EGSRLPILNRSTGTRKEHSHSSLPVSTVRRRTGSSSTLSASSESSERAKGEDEKHVN 2593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 75 SIDHDMNNKFKSGELKSMFNQ-----CKVEEMDF 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 2594 SI-----SGTKQSKENQVSAGKGTWRKIKENEF 2620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-08-449-731-2
; Sequence 2, Application US/08449731
; Patent No. 6413727
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
              ANAND, RAKESH
              CARLSON, MARY
              GRODEN, JOANNA
              HEDGE, PHILIP J.
              JOSLYN, GEOFF
              KINZLER, KENNETH
              MARKHAM, ALEXANDER F.
              NAKAMURA, YUSUKE
              THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
                      GENE IN COLORECTAL CANCER IN HUMANS

```

```

NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731

```

```

US-08-449-731-2
Query Match 12.7%; Score 68; DB 4; Length 2843;
Best Local Similarity 27.8%; Pred. No. 44;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAFERRVPLPMAEVPMSTONSMGSSA---SASASSLENWVKDLERLN 74
DB 2534 EPSRLPINSRGTWKREHSHKSSLPVSTWRTGSSSTLSASSESEKASEDEKHVN 2593
QY 75 SIDHDNNNKGSGELKSMFNQ-----GKVEEMDF 104
DB 2594 SI-----SGTKSQSKENQVSAKGTWRKIRENEF 2620

RESULT 14
US-08-821-355A-7
; Sequence 7, Application US/08821355A
; Patent No. 5851775
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hammer & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,355A
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 5851775E
; US-08-821-355A-7

```

QY 75 SIDHNNKFGSGELKSMFNO-----GKVEEMDF 104
 Db 2594 SI-----SGTKQSKENOVSAKGTWRKIKENE 2620

RESULT 15
 US-09-003-687A-7
 ; Sequence 7, Application US/09003687A
 ; Patent No. 5998600
 ; GENERAL INFORMATION:
 ; APPLICANT: Barker, Nick
 ; APPLICANT: Clevers, Hans
 ; APPLICANT: Korinek, Vladimir
 ; APPLICANT: Morin, Patrice
 ; APPLICANT: Kinzler, Kenneth
 ; APPLICANT: Vogelstein, Bert
 ; APPLICANT: Sparks, Andrew
 ; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
 ; TITLE OF INVENTION: Interact to Prevent Cancer
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner & Witcoff, Ltd.
 ; STREET: 1001 G Street, N.W.
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20001

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/003,687A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/821,355
 FILING DATE: 20-MAR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A
 REGISTRATION NUMBER: 32,145
 REFERENCE/DOCKET NUMBER: 1107.05064
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 TELEX: 97430 BMB UT

INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2973 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: No. 5998600e
 US-09-003-687A-7

Query Match 12.7%; Score 68; DB 2; Length 2973;
 Best Local Similarity 27.8%; Pred. No. 47;
 Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRTVTMAEFRVRPLPPMAEVPMLSTONSGSSA---SASASSLLEMMWEKDLLERLN 74
 Db 2534 EFSRPLPNRSTWRKREHSHSSLLPRVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2593

QY 75 SIDHNNKFGSGELKSMFNO-----GKVEEMDF 104
 Db 2594 SI-----SGTKQSKENOVSAKGTWRKIKENE 2620

Search completed: July 16, 2003, 12:45:44
 Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 11:53:06 ; Search time 72.5 seconds
(without alignments)
4256.668 Million cell updates/sec

Title: US-09-814-661a-1

Perfect score: 2069
Sequence: 1 aatggaacacgtgtcaaca.....cctctctgtatcaatgcttt 1158

Scoring table:

BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DRV=xlp
-O=/cnp2_1/USPTO.spool/US09814661/runat_16072003_115257_14506/app_query.fasta_1.1351
-DB=A_Geneseq_101002 -qfmt=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=pct -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US09814661@cgn_1.114 @runat_16072003_115257_14506 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	537	26.0	104	21	AA990470	Yeast Sm11 (suppre
2	103	5.0	139	22	AA005979	Human polypeptide
3	101	4.9	307	21	AA181310	Plasmodium falcip
4	100	4.8	319	22	ABG19708	Novel human diagn
5	94.5	4.6	885	23	ABP35691	Fungal ZBC protein
6	93	4.5	161	22	AA005940	Human polypeptide
7	93	4.5	1080	22	AA000803	Human bone marrow
8	93	4.5	1637	22	AA000916	Human bone marrow
9	93	4.5	2263	22	AA079000	Human protein SEQ
10	91	4.4	126	22	AA000187	Human polypeptide
11	90	4.3	525	22	AAU12199	Human 22438 sulfat
12	90	4.3	525	22	AA085480	Human 22438 sulfat
13	88.5	4.3	157	23	AA084207	Human cytochrome b
14	88.5	4.3	321	22	AA072655	Murine OR-like pol
15	88.5	4.3	322	22	AAU30371	Novel human secret
16	88.5	4.3	552	22	AB112445	Human bone marrow
17	88.5	4.3	1186	20	AA05839	Banana ripening fr
18	88	4.3	2016	22	AB071487	Drosophila melanog
19	87.5	4.2	86	22	AA075549	Human colon cancer
20	87	4.2	68	22	AA009397	Human polypeptide
21	87	4.2	216	23	AAU69527	Human G protein-co
22	86	4.2	494	22	ABG29246	Novel human diagn
23	85.5	4.2	351	21	AA048481	Arabidopsis thalia
24	85.5	4.2	1037	21	AA050663	Arabidopsis thalia
25	85.5	4.2	1104	21	AA050662	Arabidopsis thalia
26	85.5	4.2	1148	21	AA050661	Arabidopsis thalia
27	85	4.1	372	23	AB049715	Listeria monocytog
28	84.5	4.1	78	22	AA013043	Human polypeptide
29	84.5	4.1	324	22	AA072653	Murine OR-like pol
30	84.5	4.1	921	18	AA022480	Plasmodium falcip
31	84.5	4.1	921	21	AA077902	P. falciparum ebl-
32	84.5	4.1	1665	22	AB064010	Drosophila melanog
33	84	4.1	167	23	AA051969	Human mitochondria
34	84	4.1	1982	22	ABG16404	Novel human diagn
35	84	4.1	1982	22	ABG16404	Novel human diagn
36	83	4.0	103	23	AB089068	Human polypeptide
37	83	4.0	109	22	AA013227	Human polypeptide
38	83	4.0	567	21	AA024967	Plant SDF encoded
39	83	4.0	799	21	AA024966	Plant SDF encoded
40	83	4.0	817	21	AA024965	Plant SDF encoded
41	82.5	4.0	94	22	AA091044	Human immune/haema
42	82.5	4.0	318	22	AA010881	G-protein coupled
43	82.5	4.0	319	22	AA010880	G-protein coupled
44	82.5	4.0	351	21	AA026997	Arabidopsis thalia
45	82.5	4.0	1187	20	AA05841	Banana ripening fr

ALIGNMENTS

RESULT 1
AA990470
ID AA990470 standard; Protein; 104 AA.
XX
XX
AC AA990470;
XX
XX 15-AUG-2000 (first entry)
DT
DT Yeast Sm11 (suppressor of mecl lethality) protein.
DE
DE
XX
XX Sm11 protein; YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT.
XX
XX Saccharomyces cerevisiae.
XX
XX W0200017225-A2.

XX 30-MAR-2000.
 PD
 PF 24-SEP-1999; 99WO-US22260.
 PR 24-SEP-1998; 98US-0158858.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Rothstein R, Zhao X;
 XX WPI: 2000-283539/24.
 DR N-PSDB; AAA14263.
 PT New Smll protein and its homologs, useful for treating cancer,
 PT microbial infection and ataxia telangiectasia and in screening for
 PT specific modulators
 XX
 PS Claim 2; Fig 1C; 98pp; English.
 XX This sequence represents the yeast Smll (suppressor of mecl lethality)
 CC protein. This protein is encoded by the SML1 gene, located on chromosome
 CC XIII, and is an inhibitor of ribonucleotide reductase. The SML1 gene
 CC (assigned the name YML058w by the Stanford yeast genomic project) was
 CC initially isolated in a mecl-1 strain by genetic analysis. Smll is a
 CC suppressor of the mecl mutant in yeast which is associated with abnormal
 CC levels of recombination in both meiosis and mitosis. Smll also permits
 CC cell growth in the absence of the cell cycle checkpoint proteins Mecl
 CC and Rad53, which, unlike most checkpoint proteins, are essential for cell
 CC growth. Smll binds to the largest subunit of ribonucleotide reductase
 CC (RNR) which catalyses the rate-limiting step of deoxynucleotide
 CC triphosphate (dNTP) synthesis, conversion of NTPs to dNTPs. Binding of
 CC Smll reduces the activity of RNR and thus inhibits the synthesis of
 CC dNTPs. The SML1 gene, or fragments thereof may be used to isolate the
 CC human homologue of SML1. The Smll protein, and compounds that modulate
 CC the interaction of Smll with ribonucleotide reductase (RNR), may be used
 CC to alter the rate at which cells divide. These are particularly useful
 CC for treating cancer, microbial infection and ataxia telangiectasia (AT),
 CC a condition in which patients are prone to cancer. Smll may also be used
 CC to screen for modulatory agents, to raise specific antibodies, and for
 CC stimulating the function of the ATM gene (a mammalian Mecl homologue
 CC which is mutated in AT). Anti-Smll antibodies are used as diagnostic and
 CC analytical immunoassay reagents and to remove Smll from serum or to
 CC titrate Smll intracellularly.
 XX
 SQ Sequence 104 AA;

Alignment Scores:
 Pred. No.: 5.91e-52 Length: 104
 Score: 537.00 Matches: 104
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.93% Indels: 0
 DB: 21 Gaps: 0

US-09-814-661A-1 (1-1158) x AAY90470 (1-104)

QY 549 ATGCAAAATCCAGAGACTACTTTTACGCTCAAAATCGCTGCCAACAACAAGCCCT 608
 Db 1 MetGlnAsnSerGlnAspPyrPheThrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 QY 609 TCCACATTCGCTACCGCTACCATGCGGGAATTTAGAAAGGTGCCTTTGCCACCTATGGCT 668
 Db 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuProMetAla 40
 QY 669 GAGTTTCTATTTGTTCTACTCAAACTCCATCGGAGCTCGCTTTCCTCCCTCTCT 728
 Db 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
 QY 729 TCATTAGAAATCGGGAAGGATTTGGAGGAGAGACTCACTCTATCGATCATCATG 788
 Db 61 SerLeuGluMetTrpGlnAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80

QY 789 AACACAACAAATTTGGTTCTGGCGAAGCTAAATATATATGTTCAACAGGGTAAGTCGAG 848
 Db 81 AsnAsnAsnLysPheGlySerGlyGlyLeuLysSerMetPheAsnGlnGlnLysValGlu 100
 QY 849 GAAATGGACTTC 860
 Db 101 GluMetAspPhe 104
 RESULT 2
 AAO05979
 ID AAO05979 standard; Protein; 139 AA.
 XX
 AC AAO05979;
 XX
 XX 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 19871.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 OS Homo sapiens.
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 XX 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-514838/56.
 DR N-PSDB; AAI85910.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 19871; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 139 AA;
 Alignment Scores:
 Pred. No.: 0.0121 Length: 139
 Score: 103.00 Matches: 28
 Percent Similarity: 48.68% Conservative: 9
 Best Local Similarity: 36.84% Mismatches: 33
 Query Match: 4.98% Indels: 6
 DB: 22 Gaps: 2

US-09-814-661A-1 (1-1158) x AAO05979 (1-139)

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridization probes, CC polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome CC mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABG0010-ABG0377 represent novel human CC diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 319 AA:

Alignment Scores:
 Pred. No.: 0.0364 Length: 319
 Score: 100.00 Matches: 70
 Percent Similarity: 38.10% Conservativity: 34
 Best Local Similarity: 25.64% Mismatches: 93
 Query Match: 4.83% Indels: 76
 DB: 22 Gaps: 16

US-09-814-661A-1 (1-1158) x ABG19708 (1-319)

QY 389 TTTTGGACATCGCGGTTTCGCCCGATTCGCTCAGCGGGTAT----- 433
 ||||| :|:|:|:| | | | :|:
 Db 66 PheTrp---ValSerGluPheLeuHisGlnThrGlyThrTyrlleHisIleTyrlleu 84
 QY 434 -----AAAAAGAGATCTTTTTCCTGCTGCTCCCTCCATTTTAAATGCTT 484
 ||||| :|:|:|:| | | | :|:
 Db 85 LeuPheIleSerMetLysSerPhePheLeuTyrlleAlaSerPro----- 98
 QY 485 ATCTGCTCTTTGTGATCTTACGGTCTCACTAACCTCTCTTCAACTGCTCAATAATTTC 544
 :|:|:|:|:|:|:| | | | :|:
 Db 99 ---LeuThrPheMetIleLeu---SerHis-----PheGlnAlaSerSer 113
 QY 545 CGCATGCAAAATCCCAAGACTATTTCAGCTCAAAATCGCTGCCAACAACAACAGC 604
 :|:|:|:| | | | :|:
 Db 114 IlePheGly---PheCysSerSerLeuPheArgGlnSerPhePheSerAsnIleLeu 132
 QY 605 CCCTTCCACATTCGCTACCGTACCATCGCGGAATTTAGA-----AGGTGCC 652
 ||| :|:|:|:| | | | :|:
 Db 133 PhePhe***Val-TyrThrGlnSerTyrllePheArgGlnThrTyrllePhe 152
 QY 653 TTGCGCACTATGCTGAGGTC-----CTATGTTGCTACTCAAACTCCAT 700
 | | | :|:|:|:| | | | :|:
 Db 152 o-CysGlnPheAspIleThrPheValIlePheLysMetCysCysIleLeu----- 168
 QY 701 GGGCAGCTCCGCTTCGCTCGCTTCATTAGAAATGGGAAAGGATTGGAGGA 760
 :|:|:|:| | | | :|:
 Db 169 -----IleLeu-----CysPheAspLeuLeuAsnArg 178
 QY 761 GAGACTCAACTCTATCGATCATGACATGAACAACAACAAATTTGCTGCGCAACTAAA 820
 :|:|:|:| | | | :|:
 Db 178 rgAsnGlySerIleIleIleGlnLysPheGlyIleTyrlleGlyLeuValLysIlePheIleP 198
 QY 821 ATCTATGTTCAACAGGGTAGGTCGAGGAAGTCTTCTAAAGTCTCTTCTTCACTC- 879
 ||| :|:|:|:| | | | :|:
 Db 198 he***IlePheThrAsnIleHisLeuLeuGln***PhePheGlnValPheIleLeuG 218
 QY 880 -----TTTCTTTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 919
 |||||:|:|:|:| | | | :|:

Db 218 lnPhePheAsnIleValLysLysSerPheLeuGlnProAlaGlySerLysArg***LysP 238
 QY 920 TCTTCTCTTAGATACCCCTCTTTTTCAGGAGACTCTCGCTCTACTATTTCTGTCATTCGCA 979
 |||||:|:|:|:| | | | :|:
 Db 238 hePheSerLysValArgPhePheSer-----PheLeu***Phe***L 252
 QY 980 AA-----CATTTCTCTCCCGTGCATTTTTCCTTTTCCCTTATATATATATATAT- 1026
 ||| :|:|:|:| | | | :|:
 Db 252 ySTpGluAlaArgSerSerAlaSerLysLeuPheLeuLeuIleTyrlle***ValTyrl 272
 QY 1027 -----ATATATATATATATATATCT-----CTCT 1048
 :|:|:|:| | | | :|:
 Db 272 rgIlePheSerPheLeuValPheProValValCysProValGlyTrpTyrlleLeuP 292
 QY 1049 TCTACGTATTTTCTAT---TTCGTCTCTTTATC 1080
 |||||:|:|:|:| | | | :|:
 Db 292 hePhePhePheLeuTyrllePhePheLeuPheLeu 303
 RESULT 5
 ABP35691
 ID ABP35691 standard; Protein; 885 AA.
 XX AC ABP35691;
 XX 24-JUL-2002 (first entry)
 XX Fungal ZBC protein sequence #117.
 XX Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;
 KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;
 KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;
 KW angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
 KW fungal toxin; cell surface receptor; plant growth regulator; pigment;
 KW insecticide; antineoplastic.
 XX Unidentified.
 OS WO200224865-A2.
 XX PN 28-MAR-2002.
 XX PD 19-SEP-2001; 2001WO-US29288.
 XX PF 19-SEP-2000; 2000US-233564P.
 XX PR (MICR-) MICROBIA INC.
 XX PA Holtzman D, Madden K, Maxon M, Sherman A;
 XX WPI; 2002-352005/38.
 XX DR N-PSDB; ABN79880.
 XX PT New method for improving the production of a secondary metabolite e.g.
 PT antineoplastic agent, ergot alkaloid from a fungus involves modulation
 PT of the expression of at least one zinc binuclear cluster protein gene
 PS Disclosure; SEQ ID 246; 49pp + sequence listing; English.
 XX The invention relates to improving the production of a secondary
 CC metabolite by a fungus. This involves modulating the expression of at
 CC least one ZBC (zinc binuclear cluster protein) gene in a manner to
 CC improve the yield of the secondary metabolite. Methods of the invention
 CC may be used for improving the production of the secondary metabolite e.g.
 CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such
 CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),
 CC as ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such
 CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,
 CC a fungal toxin, a modulator of cell surface receptor signalling, a plant
 CC growth regulator, a pigment, an insecticide, or an antineoplastic
 CC compound. The method results in a decrease in fermentor run-time, a
 CC decrease in the size of the fermentor required for the production of
 CC equivalent amounts of the secondary metabolite, or a decrease in the


```
Db 194 rleuSerSerLeuProGlnAspLysIlePheSerProSerValThrSer----- 211
QY 512 CACTAACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATCCCAAGACTACTT 571
Db 212 -ValAsnThrMetThr-----AlaPheSerThrSerGlnAsnGlySe 226
QY 572 TTACCTCAAAATCGCTGCCAACACAAACAGCCCTTCCACATGCGTACCGTGACCAT 631
Db 226 rleuSerGlnSerValSerGln-----ProThrThr----- 237
QY 632 GCGCGAATTAGAGGGTGCCTTTGCCACCTATGCT-----GAGGTTCCTATCTT 682
Db 238 -----GluGlyAlaProPocysGlyLeuAsnLysGluGlnSerAsnLe 252
QY 683 GTCTACTCAAACTCCATGGG-----AGCTCCGCTTCTCGCTCCGCTTCTTC 730
Db 252 uleuProAsnSerLeuLysValPheAsnPheAsnSerSerThrSerHisSerSe 272
QY 731 ATTA-----GAAATGTGGAAAGGATT 754
Db 272 rleuLysSerProSerHisMetGluLysTyProGlnLysGluLysThrLysGluAspLe 292
QY 755 GGAGGAGAGACTCAAC-----TCTAT 775
Db 292 uaspSerArgSerAsnLeuHisLeuProGluThrLysPheSerGluLeuSerLysLeuLy 312
QY 776 CGATCATGACATGACAAACAACAATTTGGTCTGCGCAACTTAAATCTATGTTCAACCA 835
Db 312 sAsnAspMetGluLysAlaAsnHisIleGluSerValIleLysSerAsnLeuProAs 332
QY 836 GGTAAAGTCGAGGAATGAGCTTCTAAAGTCTCTTCACTCTTCTTCTTCTTCTTCTTCT 895
Db 332 nCysAlaAsnSerAspThrAspPheMetGlyLeuPheLysSerSerArgTyAspProSe 352
QY 896 CATTTCCTCCACTAGTCTGCTCTTCTCTCTCTAGATACCTCTTCTTCAGGAGCTCn-- 953
Db 352 rleuSer-----PheSerGlyMetSerLeuSerAspThrMetThrLeuArgGlySerVa 370
QY 954 -----CGTCTACTATTGTTCTCAT----- 974
Db 370 lGlnAsnLysLeuAsnProArgProGlyLysValValIleTySerGluProAspValSe 390
QY 975 -----CTCGAAGACTTCTCTCCCGTG 995
Db 390 rGluLysCysIleGluValPheSerAspIle 400
```

RESULT 8

```
AAM00916
ID AAM00916 standard; Protein; 1637 AA.
XX AC AAM00916;
XX DT 01-OCT-2001 (first entry)
XX DE Human bone marrow protein, SEQ ID NO: 392.
XX KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
XX OS Homo sapiens.
XX PN WO200153453-A2.
XX PD 26-JUL-2001.
XX PF 23-DEC-2000; 2000WO-US34960.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 30-NOV-2000; 2000US-0250583.
XX (HYSE-) HYSEQ INC.
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
XX Ren F, Wang J, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Drmanac RT;
XX WPI; 2001-488707/53.
XX N-PSDB; AAH90035.
DR Novel bone-marrow-expressed polynucleotides and polypeptides, useful
DR for treating e.g. cancer and immune deficiency disorders -
XX Claim 10; Page 477-481; 648pp; English.
XX The present sequence is one of 251 novel human polypeptides encoded
XX by a bone marrow-expressed polynucleotide. The polynucleotide and the
XX polypeptide encoded by it are useful in the treatment of various
XX immune deficiencies and disorders. The deficiencies and disorders may
XX be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
XX infection, or may result from an autoimmune disorder, a coagulation
XX disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
XX suppression of an inflammatory response or treatment of a nervous
XX system disorder such as Alzheimer's disease. Detection of the presence
XX or increased expression of the polynucleotide or the protein it
XX encodes is useful for the diagnosis and/or prognosis of one
XX or more types of cancer. The polynucleotide and polypeptide can be
XX used as nutritional sources or supplements and in the screening of
XX chemical compounds as potential drugs.
XX SQ Sequence 1637 AA;

Alignment Scores:
Pred. No.: 0.426 Length: 1637
Score: 93.00 Matches: 74
Percent Similarity: 35.05% Conservative: 28
Best Local Similarity: 25.43% Mismatches: 108
Query Match: 4.49% Indels: 81
DB: 22 Gaps: 13

US-09-814-661A-1 (1-1158) x AAM00916 (1-1637)
QY 293 CCGCTGAGCCCAACAGGGCTCCACTACCGCCGGTGGCCATTTTGGGAAGTCATCGG 352
Db 692 LysLeuArgProLysArgAlaSerAlaGluGlnSerValLeuPheLysSerLeuHisThr 711
QY 353 TCCCAAAAAGGAATAGCCATACATATCGTTACTGTTTGGCAACATCCCGGTTTCGCC 412
Db 712 AsnThrAsnGlyAsnSerGluProLeuValMetProGluIleAsn-AspLysGluAsnAr 731
QY 413 CGATTCCCGCTCAGCGGGTATATAAAGAGAT-----CTTTTTCCTTC 454
Db 731 gAspValThrAsnGlyGlyIleLysArgSerArgLeuGluLysSerAlaLeuPheSerSe 751
QY 455 CTGGGTG---TCCCTTCATTTTAAATGCTTATCTGCTCTCTTGTGATCTTACGGTCT 511
Db 751 rLeuLeuSerSerLeuProGlnAspLysIlePheSerProSerValThrSer----- 768
QY 512 CACTAACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATCCCAAGACTACTT 571
Db 769 -ValAsnThrMetThrThr-----AlaPheSerThrSerGlnAsnGlySe 783
QY 572 TTACGCTCAAAATCGCTGCCAACACAAACAGCCCTTCCACATGCGTACCGTGACCAT 631
Db 783 rleuSerGlnSerSerValSerGln-----ProThrThr----- 794
QY 632 GCGCGAATTAGAGGGTGCCTTTGCCACCTATGCT-----GAGGTTCCTATGTT 682
```

Db 795 -----GluGlyAlaProCysGlyLeuAsnLysGluGlnSerAsnLe 809
|||
Qy 683 GTCTACTAAACCTCCATGGC-----AGCTCCGCTTCTGCTCCGCTTCTTC 730
|||
Db 809 uLeuProAspAsnSerLeuLysValPheAsnSerSerThrSerHisSerSe 829
|||
Qy 731 ATTA-----GAAATGTGGAAAGGATT 754
|||
Db 829 rLeuLysSerProSerHisMetGluLysTyrProGlnLysGluLysThrLysGluAspLe 849
|||
Qy 755 GGAGAGAGACTCAAC-----TCAT 775
|||
Db 849 uAspSerArgSerAsnLeuHisLeuProGluThrLysPheSerGluLeuSerLysLeuLy 869
|||
Qy 776 CGATCATCATGACATGAACAACAATTTGGTCTGGCGAACAATAATATATGTTCAACCA 835
|||
Db 869 sAsnAspAspMetGluLysAlaAsnHisIleGluSerValLysSerAsnLeuProAs 889
|||
Qy 836 GGGTAAGTCGAGGAATGGACTTCTAAAGTTCTTTCATATCTTCTTCTTCTTCTTCTTC 895
|||
Db 889 nCysAlaAsnSerAspThrAspPheMetGlyLeuPheLysSerSerArgTyrAspProSe 909
|||
Qy 896 CATTTCCCACTAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC 953
|||
Db 909 rIleSer-----PheSerGlyMetSerLeuSerAspThrMetThrLeuArgGlySerVa 927
|||
Qy 954 -----CGTCTACTATTGTTGTCATT----- 974
|||
Db 927 lGlnAsnLysLeuAsnProArgProGlyLysValLysValLysValLysValLysValLys 947
|||
Qy 975 -----CTCGAAACATTCTCTCCCGTG 995
|||
Db 947 rGluLysCysIleGluValPheSerAspIle 957
|||

RESULT 9

AAW79000

ID AAW79000 standard; Protein: 2263 AA.

XX

AC AAW79000;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human protein SEQ ID NO 1662.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200157190-A2.

XX

PD 09-AUG-2001.

XX

PF 05-FEB-2001; 2001WO-US04098.

XX

PR 03-FEB-2000; 2000US-0496914.

XX

PR 27-APR-2000; 2000US-0560875.

XX

PR 20-JUN-2000; 2000US-0598075.

XX

PR 19-JUL-2000; 2000US-0620325.

XX

PR 01-SEP-2000; 2000US-0654936.

XX

PR 15-SEP-2000; 2000US-0663561.

XX

PR 20-OCT-2000; 2000US-0693325.

XX

PR 30-NOV-2000; 2000US-0728422.

XX

DA (HYSE-) HYSEQ INC.

XX

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;

WPI; 2001-476283/51.

DR N-PSDB; AAK52133.

XX

PT Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

XX

PS Claim 20; Page 4006-4010; 6221pp; English.

XX

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

CC (AAW80020) are omitted as the relevant pages from the sequence listing

CC were missing at the time of publication.

XX

SQ Sequence 2263 AA;

Alignment Scores:

Pred. No.: 0.482 Length: 2263

Score: 93.00 Matches: 74

Percent Similarity: 35.05% Conservatives: 28

Best Local Similarity: 25.43% Mismatches: 108

Query Match: 4.49% Indels: 81

DB: 22 Gaps: 13

US-09-814-661a-1 (1-1158) x AAW79000 (1-2263)

Qy 293 CGCGCTGAGCCAAACGGCTCCACTACCCCGCGCGGTGCCATTTTGGGAAGTCATCCG 352

Db 1318 LysLeuArgProLysArgAlaSerAlaGluGlnSerValLeuPheLysSerLeuHisThr 1337

Qy 353 TCCCAAAAGGAATAGCCATATCATCTGTTACTGTTTGGACATCGCCGCTTCCGCC 412

Db 1338 AsnThrAsnGlyAsnSerGluProLeuValMetProGluIleAsn-AspLysGluAsnAr 1357

Qy 413 CGATTCCGCTCAGCGGTATATAAAGAGAT-----CTTTTCTTTTTC 454

Db 1357 gAspValThrAsnGlyGlyIleLysArgSerArgLeuGluLysSerAlaLeuPheSerSe 1377

Qy 455 CTGCGTG---TCCCTTCATTTTAAATGCTTATCTGCTCTCTTGTGATCTTACGGTCT 511

Db 1377 rLeuLeuSerSerLeuProGlnAspLysIlePheSerProSerValThrSer----- 1394

Qy 512 CACTAACCTCTCTTCAACTGCTCAATATTTCCGCTATGCAAAATTTCCCAAGACTACTT 571

Db 1395 -ValAsnThrMetThrThr-----AlaPheSerThrSerGlnAsnGlySe 1409

Qy 572 TTACGCTCAAAATCGCTGCCAACCAACAGCCCTTCCACATTCGCTACCGTGACCAT 631

Db 1409 rLeuSerGlnSerSerValSerGln-----ProThrThr----- 1420

Qy 632 GCGGGAATTTAGAAAGGTGCGCTTTGGCCACCTATGCGT-----GAGGTTCCTATGTT 682

Db 1421 -----GluGlyAlaProProCysGlyLeuAsnLysGluGlnSerAsnLe 1435

Qy 683 GTCTACTCAAAATCCCATGGGC-----AGCTCCGCTTCTCCCTCCGCTTCTTCTTC 730

Db 1435 uLeuProAspAsnSerLeuLysValPheAsnPheAsnSerSerThrSerHisSerSe 1455

Qy 731 ATTA-----GAAATGTGGAAAGGATT 754

Db 1455 rLeuLysSerProSerHisMetGluLysTyrProGlnLysGluLysThrLysGluAspLe 1475

Qy 755 GGAGGAGAGACTCAAC-----TCAT 775

PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart JA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-408281/43.
 DR N-PSDB; AAS21271.
 XX
 XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical .
 XX
 PS Claim 12; Fig 56; 813pp; English.
 XX
 CC AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 525 AA;

Alignment Scores:
 Pred. No.: 0.601 Length: 525
 Score: 90.00 Matches: 73
 Percent Similarity: 34.25% Conservative: 39
 Best Local Similarity: 22.32% Mismatches: 100
 Query Match: 4.35% Indels: 115
 DB: 22 Gaps: 20

US-09-814-661A-1 (1-1158) x AAU12199 (1-525)

QY 32 ACCGGCTACTTATTCCTCC-----AAGGATCACGTT 61
 ::::|||||:|||||:|||||
 Db 16 SerGlyPheLeuTyProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
 QY 62 CTTTCTGCCAACATCATTTGGCGTCCGACGTCGCGCGTCTTCTGACATTGGT----- 115
 :||:::|||||:|||||:|||||
 Db 36 ProAsnPheValIleLeuAlaAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
 :||:::|||||:|||||:|||||
 QY 116 -----AGATACTTCCAACTAAGAG----- 136
 :||:::|||||:|||||:|||||
 Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
 :||:::|||||:|||||:|||||

QY 137 -----CATGCTTCTCTTTTGTAGGCCAATGATGAAG-----AACAA 181
 :|||:|||||:|||||
 Db 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
 QY 182 AGATTATAAATACGTACGAATATAGTAGATATGTTTATGTTTATGACCTCGTACATAGG 241
 :|||:|||||:|||||:|||||
 Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
 QY 242 AATAATTGACGCTTTTGTGGCCAAACATTTGAAATTTTGTACCTCGCGCTGAG 301
 :|||:|||||:|||||:|||||
 Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
 QY 302 CCCAAACGGCTCCACTACCGCGCGTCCCATTT----- 337
 :|||:|||||:|||||
 Db 119 ThrThrLeuAlaGluValLeuGlnGlnAlaGlyTyrValThrGlyIleLeuGlyLysTrp 138
 QY 338 ---TTGGGAAGTCATCGTCCCAAAAGAGCAATAGCATATATGTTACTGTTTGG 394
 :|||:|||||:|||||:|||||
 Db 139 HisLeuGlyHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheGlu 158
 QY 395 A-----ACATC 400
 :|||:|||||:|||||
 Db 158 yileProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
 QY 401 GCCCGTTTCGCC-----GATTCGCGCTCAGCGGTATAAAAGACATCTTTTTC 454
 :|||:|||||:|||||:|||||
 Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
 QY 455 CTGCTGTCCCTCCATTT-----TTAAATGTCTTATCTCTCTCTTGTGTCATCTTA 505
 :|||:|||||:|||||:|||||
 Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGlnProVal----- 214
 QY 506 GGGTCTCACATACTCTCTCAACTGCTCAATAATTTCCCGCTATGCCAAATCCCAAGA 565
 :|||:|||||:|||||:|||||
 Db 215 -----AsnLeuSerSerLeuAlaGln----- 221
 QY 566 CTACTTTACGCTCAAAATCGCTCCCAACAACAAGCCCTTCCACA----- 614
 :|||:|||||:|||||:|||||
 Db 222 ---LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
 QY 615 ---TTGCGTACCGTGACCATGCGGAATTTAGAAAGGTGCGCTTTGCCACCTATGGCTGA 670
 :|||:|||||:|||||:|||||
 Db 240 oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu---ProValThrGlu 258
 QY 671 GGTTCCTATGTTGCTACTCAAAACTCCATGGGAGCTCCGCTTCTCCCTCCGCTTCTTC 730
 :|||:|||||:|||||:|||||
 Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly----- 272
 QY 731 ATTAGAATGTGGAAAGGAT-----TTGGAGGAGAGACTCAACTCTATCGATCATGA 784
 :|||:|||||:|||||:|||||
 Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHisTh 290
 QY 785 CATGAACAACAACAATTT 803
 :|||:|||||:|||||
 Db 290 rValLysGluAsnThrPhe 296
 RESULT 12
 AAB5480
 ID AAB5480 standard; Protein; 525 AA.
 XX
 AC AAB5480;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE Human 22438 sulfatase polypeptide.
 XX
 KW Sulfatase; 22438 sulfatase; nootropic; neuroprotective; antibacterial;
 KW antiinflammatory; vasotropic; antitumor; gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200155411-A2.

XX PD 02-AUG-2001.
 XX PF 31-JAN-2001; 2001WO-US03266.
 XX PR 31-JAN-2000; 2000US-0495823.
 XX PA (MILL-) MILLENNIUM PHARM INC.
 XX PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
 XX PR WPI; 2001-476214/51.
 XX DR N-PSDB; AAH46860, AAH46864.
 XX PT Novel human sulfatase polypeptides useful for treating and diagnosing
 XX PT sulfatase-related disorders such as cerebrovascular diseases, acute
 XX PT meningitis, multiple sclerosis, degenerative diseases and tumor
 XX PS Claim 9; Fig 1; 180pp; English.
 XX CC The invention provides 22438, 23553, 25278 or 26212 human sulfatase
 CC polypeptides and polynucleotides. The sulfatase genes and polypeptides
 CC are useful for treating disorders involving the brain such as
 CC cerebrovascular diseases, infections such as acute meningitis,
 CC demyelinating diseases including multiple sclerosis, degenerative
 CC diseases affecting the cerebral cortex including Alzheimer's disease
 CC and Pick disease, spinocerebellar degenerations including spinocerebellar
 CC ataxias including Friedreich ataxia, and ataxia telangiectasia,
 CC degenerative diseases affecting motor neurons including amyotrophic
 CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
 CC toxic and acquired metabolic diseases, including vitamin deficiencies,
 CC and neurocutaneous syndromes (phakomatosis) including neurofibromatosis.
 CC The present sequence represents a human 22438 sulfatase polypeptide.
 XX SQ Sequence 525 AA;

Alignment Scores:
 Pred. No.: 0.601 Length: 525
 Score: 90.00 Matches: 73
 Percent Similarity: 34.25% Conservative: 39
 Best Local Similarity: 22.32% Mismatches: 100
 Query Match: 4.35% Indels: 115
 DB: 22 Gaps: 20

US-09-814-661A-1 (1-1158) x AAB85480 (1-525)

QY 32 ACGGCTACTTATTCCTCC-----AAGGATCAGTT 61
 Db 16 SerGlyPheLeuTyrProLeuValAspPheCysIleSerGlyLysThrArgGlyGlnLys 35
 QY 62 CTTCTGCCAACATCATCGCTCGAACGTCGCGGCTCTTCTGCACATGCT-----115
 Db 36 ProAsnPheValIleIleLeuAlaAspAspMet---GlyTrpGlyAspLeuGlyAlaAsn 54
 QY 116 -----AAGAACTACTTCCAACTAAGAG-----136
 Db 55 TrpAlaGluThrLysAspThrAlaAsnLeuAspLysMetAlaSerGluGlyMetArgPhe 74
 QY 137 -----CATGCTTCTCTTTTCTTTTGTAGGCCAATGATAGAAAG-----AACAA 181
 Db 75 ValAspPheHisAlaAlaSerThrCysSerProSerArgAlaSerLeuLeuThrGly 94
 QY 182 AGATTATAATACGTCAGATATAGTAGATATGTTTTATGTTTAGACCTCGTACATAGG 241
 Db 95 ArgLeuGlyLeuArgAsnGlyValThr-----Arg 104
 QY 242 AATAATGACGTTTTTTTGTGGCCACACATTGAATTTTTTTTGTACTCGCGCTGAG 301
 Db 105 Asn-----PheAlaValThrSerValGlyGly-----LeuProLeuAsnGlu 118
 QY 302 CCCAAACGGGCTCCACTACCGCGCGGTCGCCATT-----337
 Db 119 ThrThrLeuAlaGluValLeuGlnAlaGlyTyrValThrGlyIleIleGlyLysTrp 138

QY 338 ---TTGGGAAGTCATCCGTCCCAAAAGAAATAGCCATAACATATCGTTACTGTTTGG 394
 Db 139 HisLeuGlyHisGlySerTyrHisProAsnPheArgGly-PheAspTyrTyrPheG 158
 QY 395 A-----ACATC 400
 Db 158 ylleProTyrSerHisAspMetGlyCysThrAspThrProGlyTyrAsnHisProProCy 178
 QY 401 GCCCGTTTCGCC-----GATTCGCCCTCAGGGGTATATAAAAGAGATCTTTTTC 454
 Db 178 sProAlaCysProGlnGlyAspGlyProSerArgAsnLeuGlnArgAspCysTyrThrAs 198
 QY 455 CTGGCTGTCCTTCCATT-----TTAAATGTCTTATCTGCTCTTCTGTGATCTTA 505
 Db 198 p---ValAlaLeuProLeuTyrGluAsnLeuAsnIleValGluGlnProVal-----214
 QY 506 CGGTCTCACTAACCTCTCTCAACTGCTCAATAATTTCCCGGTATGCAAAATCCCAAGA 565
 Db 215 -----AsnLeuSerSerLeuAlaGln-----221
 QY 566 CTACTTTTACGCTCAAAATCGCTGCCAACACAAACAGCCCTTCCACA-----614
 Db 222 ----LysTyrAlaGluLysAlaThrGlnPheIleGlnArgAlaSerThrSerGlyArgPr 240
 QY 615 ---TTGCGTACCGTGACCATGGCGGAATTTAGAGGGTGCCCTTGGCCACATGCTGTA 670
 Db 240 oPheLeuLeuTyrValAlaLeuAlaHisMet---HisValProLeu---ProValThrG 258
 QY 671 GGTTCCTATGTGCTCTACTCAAAACCTCCATGGCAGCTCGGCTTCTGCTCCCTCTTC 730
 Db 258 nLeuProAlaAlaProArgGlyArgSerLeuTyrGlyAlaGly-----272
 QY 731 ATTAGAAATGTGGGAAAGGAT-----TTGGAGGAGAGACTCAACTCTATCGATCATGA 784
 Db 273 -----LeuTrpGluMetAspSerLeuValGlyGlnIleLysAspLysValAspHis 290
 QY 785 CATGAACAACAACAATTT 803
 Db 290 rValLysGluAsnThrPhe 296
 RESULT 13
 ID AAM48207 standard; Protein: 157 AA.
 XX AC AAM48207;
 XX DT 15-MAR-2002 (first entry)
 XX DE Human cytochrome b5-17.
 XX DE Human; cytochrome b5-17; cancer; haemopathy; HIV infection; cytostatic;
 KW haemostatic; anti-HIV; gene therapy.
 XX OS Homo sapiens.
 XX PN CN1315355-A.
 XX PD 03-OCT-2001.
 XX PF 24-MAR-2000; 2000CN-0115109.
 XX PR 24-MAR-2000; 2000CN-0115109.
 XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
 XX PI Mao Y, Xie Y;
 XX DR WPI; 2002-056311/08.
 XX DR N-PSDB; ABA95530.
 XX PT Polypeptide-human cytochrome b5-17 and polynucleotide for coding it -


```
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-611725/70.
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
XX Claim 20; Page 284; 765pp; English.
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of polypeptide. The polypeptides are also useful for
XX identifying agents (agonists and antagonists) that bind to them. Cells
XX expressing the proteins are useful for identifying a therapeutic agent
XX for use in treatment of a pathology related to aberrant expression or
XX physiological interactions of the polypeptide. Vectors comprising
XX the nucleic acids encoding the polypeptides and cells genetically
XX engineered to express them are also useful for producing the proteins.
XX The proteins are useful in genetic vaccination, testing and
XX therapy, and can be used as nutritional supplements. They may be used to
XX increase stem cell proliferation; to regulate haematopoiesis; and in
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX immune suppression and/or stimulation; as anti-inflammatory agents; and
XX in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
XX sequences of novel human secreted proteins of the invention.
XX
XX Sequence 322 AA;
XX
Alignment Scores:
Pred. No.: 0.737 Length: 322
Score: 88.50 Matches: 78
Percent Similarity: 37.27% Conservative: 42
Best Local Similarity: 24.22% Mismatches: 115
Query Match: 4.28% Indels: 88
DB: 22 Gaps: 16
XX
US-09-814-661A-1 (1-1158) x AAU30371 (1-322)
QY 231 TCGTACATAGGAATATTCACGCTTTTGGCCCAACATTGAAATTTTTTTTCTTAC 290
DB |||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
47 SerTyrLeuLysLeuLeuThrIlePhe-----SerPheSerIleLeuPhe----- 61
QY 291 CTCGCGCTGAGCCCAACGGCTCCA-----CTACCCCGCGCGTGGCCATT 337
DB ||||| ||| ||| ||||| ||| ||||| |||
62 -----LeuSerSerAsnPheLeuLeuSerLeuTyrIleLeuProLeuAlaPheCysPhe 79
QY 338 TTGGGAAGTCATCCGTCGCCAAAGAAATAGCCATATCGTTACTGTTTGGGAAC 397
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 Leu-----AlaCysLeuPhePhePhe 86
QY 398 ATCGCCGCTTCGCGGATTCGCGCTCACGGGTATATAAAGAGATCTTTTTCCTG 457
DB |||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
87 CysPheLeuLeuPheArgPheLysPheIlePhePheLeuProLysSerMetPheLeuSer 106
QY 458 GCTGTCCCTTCCATTTTAAATGCTTATCTGCTCCTTGTGTATCTTACGGTCTACTAA 517
DB : : : ||||| ||| : : : ||| ||| ||| ||| ||| ||| ||| |||
107 SerAsnPheProIlePhe--CysIlePhePhePheAlaPhePheProTyrSerCysLeu 125
QY 518 CCTCTCTTCACTGCTCAATAATTTCCCGCTATGCAAAATTCCAAGACTACTTTACGC 577
DB |||: : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 ProValPheHisCysGlnThrPhe----TyrLeuLeuGlnIleValIleLeuIleHis 144
QY 578 TCAAAATCGTCCCAACACAAACAGCCCTTCCACATTTCGCTACCGTACCACGCGGA 637
DB ||||: : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 SerGlnLeuIle-----PheProPheSerSerAlaPheSerLysCysGly*** 160
QY 638 ATTAGAAGGGTGCCTTTGCCACTATGGCTGAGGTTCTATGTTCTCTACTCAAAACTC 697
DB : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 -----GlnGlyPhe***MetThrGln-----PheSerLeuArg 171
QY 698 CATGGG-----CAGTCGCGCTTCTGCCCTCCGCTTCTTATTAGAAATGTGGGA 745
```

```
||||| : : : ||| ||| ||| : : :
172 HisGlyPheGlnGlyLeuLeuIleThrPheGlyLeuIleSerPhe***LysIleAlaPhe 191
QY 746 AAAGGATTTGGAGGAGAGACT-----CAACTCTATCGATCATGCACAT 787
DB ||| ||| ||| : : : ||| : : : |||
192 LysLeuPhe***SerProThrThrPheLysLeuAsnPheLysLeuPheLysTyrLeuHis 211
QY 788 GAACAACAACAAATTTGGTTCTGGCGAAGCTAAATCATATGTTCAACACAGG---TAAGGT 844
DB : : : ||| ||| ||| |||
212 -----PheAsnLeuLysAlaLeuProSerLeuLeuGly 222
QY 845 CGAGGAATGGACTTCTAAAGTTCTTTCATACTCTTTTCTTCTTCTTCCATTTC--- 901
DB : : : ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
223 LeuAsnSerGlyLeuLeu***PheLeuSerHisGlnPheSerPheLysTyrIlePheTyr 242
QY 902 -----CCACTAGTTCTGTCTTTTCT 922
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 SerMetLysLysLeuHisLeuAsnLeuArgHisValLeuGluThrValLeuSerPhePro 262
QY 923 TCTCTTAGATACCTTCT-----TTTCAGGAGCTC 952
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 SerSerCysTyrSerSerLysPheValHisPheAlaLeuSerPheLeuPheProSerLeu 282
QY 953 TCGTCTACTATTGTCTCATTTCTCGAAACATTTCTCCCGTGCATTTTCCTT-----T 1006
DB ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 SerPhePheCys-PheLeuLeuPheArgPheLysPheIle-PhePheLeuArgLys 302
QY 1007 CCTTTTATATATATATATATATATATATATATATGCTCTTCTACGTATTTTCTATT 1066
DB |||: : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 erMetPheLeuSerSerAsnPheProIlePheSerIleLeuPheTyrLeuProIlePheP 322
QY 1067 TC 1068
DB ||
322 he 322
Search completed: July 16, 2003, 12:00:26
Job time : 84.5 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 11:53:10 ; Search time 21 Seconds
(without alignments)
3244.928 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 2069

Sequence: 1 aatgagcaacgtgtcaaca.....cctctcttgatcaatgtttt 1158

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp

-O/cgn2_1/USPTO_spool/US09814661/runat_16072003_115259_14566/app_query.fasta_1.1351
-DB-issued_Patents_AA -QFMT-fastan -SUFFIX-rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn.1.1.28/runat_16072003_115259_14566 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	84.5	4.1	921	2	US-08-568-459A-8
2	84.5	4.1	921	2	US-08-487-826B-8
3	84.5	4.1	921	4	US-09-210-288-8
4	81	3.9	1112	4	US-09-353-585-2
5	80	3.9	1112	4	US-09-353-585-3
6	80	3.9	2710	2	US-08-568-459A-12
7	80	3.9	2710	2	US-08-487-826B-12
8	80	3.9	2710	4	US-09-210-288-12
9	80	3.9	2938	5	PCT-US94-00198-3
10	78.5	3.8	365	2	US-08-724-974A-3
11	78.5	3.8	1130	4	US-09-442-100-4
12	78	3.8	151	4	US-09-134-001C-5595

13	75	3.6	229	3	US-08-833-488B-24	Sequence 24, Appl
14	75	3.6	253	3	US-08-833-488B-20	Sequence 20, Appl
15	75	3.7	418	4	US-09-134-001C-4149	Sequence 4149, Ap
16	74	3.6	2842	1	US-07-741-940-7	Sequence 7, Appl
17	74	3.6	2842	1	US-08-289-548A-7	Sequence 7, Appl
18	74	3.6	2842	1	US-08-452-654-7	Sequence 7, Appl
19	74	3.6	2842	4	US-08-449-731-7	Sequence 7, Appl
20	74	3.6	2843	1	US-07-741-940-2	Sequence 2, Appl
21	74	3.6	2843	1	US-08-289-548A-2	Sequence 2, Appl
22	74	3.6	2843	1	US-08-452-654-2	Sequence 2, Appl
23	74	3.6	2843	1	US-08-452-655B-2	Sequence 2, Appl
24	74	3.6	2843	1	US-08-452-655B-7	Sequence 7, Appl
25	74	3.6	2843	2	US-08-370-235A-2	Sequence 2, Appl
26	74	3.6	2843	3	US-08-450-582-7	Sequence 7, Appl
27	74	3.6	2843	3	US-08-450-582-7	Sequence 7, Appl
28	74	3.6	2843	4	US-08-449-731-2	Sequence 2, Appl
29	74	3.6	2973	2	US-08-821-355A-7	Sequence 7, Appl
30	74	3.6	2973	2	US-09-003-687A-7	Sequence 7, Appl
31	74	3.6	2973	4	US-09-136-605-7	Sequence 7, Appl
32	73.5	3.6	754	4	US-09-300-008B-2	Sequence 2, Appl
33	73	3.5	376	4	US-09-091-725-15	Sequence 15, Appl
34	72.5	3.5	283	1	US-08-118-270-78	Sequence 78, Appl
35	72.5	3.5	283	5	PCT-US93-08528-78	Sequence 78, Appl
36	72.5	3.5	431	2	US-08-712-709-5	Sequence 5, Appl
37	72.5	3.5	431	3	US-09-111-444-5	Sequence 5, Appl
38	72.5	3.5	431	4	US-09-541-228-5	Sequence 5, Appl
39	72.5	3.5	431	4	US-09-031-295-2	Sequence 2, Appl
40	72.5	3.5	531	2	US-08-724-974A-2	Sequence 2, Appl
41	72.5	3.5	555	3	US-08-968-563-34	Sequence 34, Appl
42	72.5	3.5	555	4	US-08-969-683A-34	Sequence 34, Appl
43	72.5	3.5	1402	4	US-09-004-838-11	Sequence 11, Appl
44	72.5	3.5	3287	2	US-08-477-451-7	Sequence 7, Appl
45	72	3.5	487	1	US-08-444-734A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-568-459A-8
; Sequence 8, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ORGANISM: Plasmodium falciparum
; ORGANISM: Plasmodium falciparum
US-08-568-459A-8

Alignment Scores:
Pred. No.: 0.197 Length: 921
Score: 84.50 Matches: 15
Percent Similarity: 79.41% Conservatives: 12
Best Local Similarity: 44.12% Mismatches: 2
Query Match: 4.08% Indels: 5
DB: 2 Gaps: 3

US-09-814-661A-1 (1-1158) x US-08-568-459A-8 (1-921)

QY 994 TGCATTTTCCTT-----TCCTTTAT-----ATACATATATATATATATATATATA 1041
|||||
Db 866 CysilePheilleCysValSerValtyrValCysValtyrValtyrValPheLeuTyMet 885
QY 1042 TGTCTCTTCTACGTA---TTTTTGTATTCTGTCTCTTTATC 1080
|||||
Db 886 CysValPheTyriIleTyPheilleTyriIleTyValPheille 899

RESULT 2

US-08-487-826B-8
; Sequence 8, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001C1P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-8

Alignment Scores:
Pred. No.: 0.197 Length: 921
Score: 84.50 Matches: 15
Percent Similarity: 79.41% Conservatives: 12
Best Local Similarity: 44.12% Mismatches: 2
Query Match: 4.08% Indels: 5
DB: 2 Gaps: 3

US-09-814-661A-1 (1-1158) x US-08-487-826B-8 (1-921)

QY 994 TGCATTTTCCTT-----TCCTTTAT-----ATACATATATATATATATATATATA 1041
|||||
Db 866 CysilePheilleCysValSerValtyrValCysValtyrValtyrValPheLeuTyMet 885
QY 1042 TGTCTCTTCTACGTA---TTTTTGTATTCTGTCTCTTTATC 1080
|||||
Db 886 CysValPheTyriIleTyPheilleTyriIleTyValPheille 899

RESULT 3

US-09-210-288-8
; Sequence 8, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-09-210-288-8

Alignment Scores:

Pred. No.: 0.197 Length: 921
Score: 84.50 Matches: 15
Percent Similarity: 79.41% Conservative: 12
Best Local Similarity: 44.12% Mismatches: 2
Query Match: 4.08% Indels: 5
DB: 4 Gaps: 3

US-09-814-661A-1 (1-1158) x US-09-210-288-8 (1-921)

QY 994 TGCATTTTCCTT-----TCCCTTTAT-----ATACATATATATATATATATATATATA 1041
Db 866 CysIlePheileCysValSerValTyValCysValTyValTyValPheLeuTyMet 885
QY 1042 TGTCTCTTCTACGTA---TTTTGTATTTCTGTGCTTTATC 1080
Db 886 CysValPheTyIleTyPheIleTyIleTyValPheIle 899

RESULT 4

US-09-353-585-2
; Sequence 2, Application US/09353585
; Patent No. 6287865

GENERAL INFORMATION:

APPLICANT: Dixon, Mark S
Jones, David A

TITLE OF INVENTION: Plant pathogen resistance genes and uses
thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA: US/09/353,585
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q
1/68

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 9506658.5
FILING DATE: 31-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-353-585-2

Alignment Scores:

Pred. No.: 0.616 Length: 1112
Score: 81.00 Matches: 51
Percent Similarity: 43.37% Conservative: 34
Best Local Similarity: 26.02% Mismatches: 70
Query Match: 3.91% Indels: 41
DB: 4 Gaps: 10

US-09-814-661A-1 (1-1158) x US-09-353-585-2 (1-1112)

QY 332 GCATTTTGGGAAGTCATCCGTCACAAAAGGAATAGCCATAACATATCTGTTACTGTTT 391
Db 586 AlaLeuAsnGlySerIleProAlaSerLeuGlyAsnLeuAsnAsnLeuSerMetLeuTy 605
QY 392 TGAACATCGCCCGTTTCGCCGATTCGCCCTCAGCGGTATATAAAGAGATCTTTTTT 451
Db 606 -----LeuTyAsnAsnGlnLeuSerGlySerIleProGluGluLeuGly 620
QY 452 TTCCTGGCTGCTCCCTTCCATTTTAAATGCTTATCTGCTCTTGTGATCTTACGGTCT 511
Db 621 TyrLeuSer-SerLeuThrTyLeuSerLeuGlyAsnAsnSerLeu-----AsnGlyLe 638
QY 512 CACTAACCTCTCTTCAACTGCTCAATAATTTCCTCGCTATGCAAAATTCACAGACTPACT 571
Db 638 uileProAlaSer-----PheAlaAsnMetArgAsnLeuGlnAlaLeu1 653
QY 572 TTACGCTCAAAATCGCTGCCAACACACACAGCCCTTCCACATTCGCTACCGTGACCAT 631
Db 653 eLeuAsnAspAsnAsnLeu---IleGlyGluIleProSerSerValCysAsnLeuThrSe 672
QY 632 GCGGAATTAGAGGGTGCCT-----PheAlaAsnMetArgAsnLeuGlnAlaLeu1 653
Db 672 rleuGluValLeuTyMetProArgAsnAsnLeuTyGlyValProGlnCysLeuG1 692
QY 659 ACCTATGCTGAGTTCCTTATGTTGCT---ACTCAAAACTCCATGGGAGCTCCGCTTC 715
Db 692 yAsnIleSerAsnLeuGlnValLeuSerMetSerAsnSerPheSerGlyGluLeuPr 712
QY 716 TCCTCCGCTTCTTCA-----TTAGAAATGTGGGAAAGGATTTGGA 757
Db 712 oSerSerIleSerAsnLeuThrSerLeuGlnIleLeuAspPheGlyArgAsnAsnLeuG1 732
QY 758 G-----GAGAGACTCAACTCTATCGAT---CATGACATGAACAA 793
Db 732 uGlyAlaIleProGlnCysPheGlyAsnIleSerSerLeuGluValPheAspMetGlnAs 752
QY 794 CAACAAATTTGTTCTGCGCAACTAAATCTATGTTCAACACAGGT 839
Db 752 nAsnLysLeu---SerGlyThrLeuProThrAsnPheSerIleGly 766

RESULT 5

US-09-353-585-3
; Sequence 3, Application US/09353585
; Patent No. 6287865

GENERAL INFORMATION:

APPLICANT: Dixon, Mark S
Jones, David A
Jones, Jonathan DG

TITLE OF INVENTION: Plant pathogen resistance genes and uses
thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/353,585
 FILING DATE: 15-Jul-1999
 CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/930,277
 FILING DATE: 27-OCT-1997
 APPLICATION NUMBER: PCT/GB96/00785
 FILING DATE: 01-APR-1996
 APPLICATION NUMBER: GB 9506658.5
 FILING DATE: 31-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ms Mary J Wilson
 REGISTRATION NUMBER: 32,955
 REFERENCE/DOCKET NUMBER: 620-69
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1112 amino acids
 TYPE: amino acid
 STRANDEDNESS: <Unknown>
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Tomato
 STRAIN: Cf2
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 US-09-353-585-3

Alignment Scores:
 Pred. No.: 0.826 Length: 1112
 Score: 80.00 Matches: 51
 Percent Similarity: 43.37% Conservative: 34
 Best Local Similarity: 26.02% Mismatches: 70
 Query Match: 3.87% Indels: 41
 DB: 4 Gaps: 10

US-09-814-661a-1 (1-1158) x US-09-353-585-3 (1-1112)

QY 332 GCATTGGAAGTCATCCGTCACCAAGAAATACCATATCGTTACTGTTT 391
 Db 586 AlaLeuAsnGlySerIleProAlaSerLeuGlyAsnLeuAsnAsnSerMetLeuTyr 605
 QY 392 TGAACATCGCCCGTTTCGCCGATTCGCCGTCACGGGTATATAAAGAGATCTTTT 451
 Db 606 -----LeuTyrAsnAsnGlnLeuSerGlySerIleProGluGluLeGly 620
 QY 452 TTCCTGGCTGCTCCCTTCCATTTTAAATGCTTATCTCTCTCTTGTGATCTACGGTCT 511
 Db 621 TyrLeuSer-IleLeuThrTyrLeuSerLeuGlyAsnSerLeu-----AsnGlyLe 638
 QY 512 CACTAACCTCTCTTCAACTGCTCAATAATATCCCGCTATGCAAAATCCCAAGACTACTT 571
 Db 638 uileProAlaSer-----PheGlyAsnMetArgAsnLeuGlnAlaLeuI 653
 QY 572 TTACGCTCAAAATCGCTCCCAACACAAAGCCCTTCCACATTCGCTACCGTACCAT 631
 Db 653 eLeuAsnAspAsnAsnLeu---IleGlyGluIleProSerSerValCysAsnLeuThrSe 672
 QY 632 GCGGAATTTAGAGGGTGCCT-----TTGCC 658
 Db 672 rLeuGluValLeuTyrMetProArgAsnAsnLeuLysGlyLysValProGlnCysLeuG 692
 QY 659 ACCATGCTGAGGTTCCTATGTGCTCT---ACTCAAACTCCATGGCGCAGCTCCGCTTC 715
 Db 692 yAsnIleSerAsnLeuGlnValLeuSerMetSerSerAsnSerPheSerGlyGluLeuPr 712

QY 716 TGCCTCCGCTCTTCTCA-----TTAGAAATGTGGAAAGGATTTCGA 757
 Db 712 oSerSerIleSerAsnLeuThrSerLeuGlnIleLeuAspPheGlyArgAsnAsnLeuG 732
 QY 758 G-----GAGAGACTCAACTCTATCGAT---CATGACATGAAACA 793
 Db 732 uGlyAlaIleProGlnCysPheGlyAsnIleSerSerLeuGluValPheAspMetGlnAs 752
 QY 794 CAACAAATTTGTTCTGCGGCAACTAAATCTATGTTCAACCCAGGT 839
 Db 752 nAsnLysLeu---SerGlyThrLeuProThrAsnPheSerIleGly 766
 RESULT 6
 US-08-568-459A-12
 : Sequence 12, Application US/08568459A
 : Patent No. 5849306
 : GENERAL INFORMATION:
 : APPLICANT: Sim, Kim L.
 : APPLICANT: Chitnls, Chetan
 : APPLICANT: Miller, Louis H.
 : APPLICANT: Peterson, David S.
 : APPLICANT: Su, Xin-zhaun
 : APPLICANT: Wellens, Thomas E.
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : AND PLASMODIUM FALCIPARUM ERYTHROCYTE-BINDING PROTEINS
 : NUMBER OF SEQUENCES: 37
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Knobbe Martens Olson & Bear
 : STREET: 620 Newport Center Drive 16th Floor
 : CITY: Newport Beach
 : STATE: California
 : COUNTRY: US
 : ZIP: 92660
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/568,459A
 : FILING DATE: 07-DEC-1995
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Israel, Ned
 : REGISTRATION NUMBER: 29,655
 : REFERENCE/DOCKET NUMBER: NIH121.001CPI
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (619) 235-8550
 : TELEFAX: (619) 235-0176
 : INFORMATION FOR SEQ ID NO: 12:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2710 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : HYPOTHETICAL: NO
 : ORIGINAL SOURCE:
 : ORGANISM: Plasmodium falciparum
 : US-08-568-459A-12

Alignment Scores:
 Pred. No.: 1.42 Length: 2710
 Score: 80.00 Matches: 14
 Percent Similarity: 57.45% Conservative: 13
 Best Local Similarity: 29.79% Mismatches: 6
 Query Match: 3.87% Indels: 14
 DB: 2 Gaps: 2

US-09-814-661a-1 (1-1158) x US-08-568-459A-12 (1-2710)

QY 994 TGCATTTTCTTCCCTTTATATACAT----- 1020
 Db 994 TGCATTTTCTTCCCTTTATATACAT----- 1020

```

Db      2664 CysValPheGlyTyrIleTyrValTyrValTyrAspPheLeuTyrMetTyrLeuTrpVal 2683
QY      1021 -----ATATATATATATATATATATATCTCTCTCTACGTATT---TTGTAATTCGT 1071
          |||||:|||||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      2684 LysAspileTyrIleTrpMetTyrLeuTyrValPheTyrIleTyrIleLeuTyrIleCys 2703
QY      1072 GTCCTTATCAAGATAGTCTA 1092
Db      2704 IleTyrIleLysLyssGuille 2710

RESULT 7
US-08-487-826B-12
; Sequence 12, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Alignment Scores:
Pred. No.: 1.42 Length: 2710
Score: 80.00 Matches: 14
Percent Similarity: 57.45% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 6
Query Match: 3.87% Indels: 14
DB: 2 Gaps: 2

US-09-814-661A-1 (1-1158) x US-08-487-826B-12 (1-2710)
QY      994 TGCATTTCTCTTCCCTTTATATACAT----- 1020
          |||||:|||:::|||||:::
Db      2664 CysValPheGlyTyrIleTyrValTyrValTyrAspPheLeuTyrMetTyrLeuTrpVal 2683

```

Db 2684 LysaspIleTyrIleTrpMetTyrLeuTyrValPheTyrIleTyrIleLeuTyrIleCys 2703
QY 1072 GTCTTTATCAAGATAGCTA 1092
Db 2704 IletyriIleLysLysGluIle 2710

RESULT 9

PCT-US94-00198-3
; Sequence 3, Application PC/TUS9400198
; GENERAL INFORMATION:
; APPLICANT: Schering Corp.
; TITLE OF INVENTION: RAS Associated GAP Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 1 Girald Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 6.0.8
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00198
; FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,824
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: DX0352 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)822-7255
; TELEFAX: (201)822-7039
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2938 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
PCT-US94-00198-3

Alignment Scores:
Pred. No.: 1.49 Length: 2938
Score: 80.00 Matches: 65
Percent Similarity: 37.81% Conservative: 56
Best Local Similarity: 20.31% Mismatches: 107
Query Match: 3.87% Indels: 92
DB: 5 Gaps: 14

US-09-814-661a-1 (1-1158) x PCT-US94-00198-3 (1-2938)

QY 312 CTCACACTACCGCCGGCTCG---CCATTTTGGGAAGTCATCCGTCACCAAGAAATA 368
Db 140 LeuTyrTyrAlaSerLysAlaPheLeuPheTrpIleMetAlaArgProLysGluTyrVal 159
QY 369 GCCATAACATATCGTTACTGTTTGGACATCGCCGTTTCGCCGGATTCGCC----- 422
Db 160 LysIle---TyrAsnAsnLeuIleSerSerAspTyrAsnSerProSerSerSerAsp 178
QY 423 -----TCAGCGGGGTATAAAAGAGATCTTTTTTTTCTCGGTGTCCTTCCATTT 473
Db 179 AsnGlyGlySerAsnAsnSerAspLysThrSerIleSerGlnLeuValSerLeuPhe 198
QY 474 TTAATGCTTTATCGTCTCTTTGTGATCTTACGGTCTCACTAACTCTCTTCAACTGCT 533

Db 199 AspAspValTyrSerThrPheSerGlySerSerLeuLeuThrAsnValAsnAsn----- 216
QY 534 CAATAAATTTCCCGCTATGCAAAATTTCCCAAGACTACTTTTAC----- 575
Db 217 -----AspHisHisTyrHisLeuHisHisSerSer 226
QY 576 GCTCAAAATCCCTGCCAACACAAACAGCCCTTCCACATTTGCGT----- 620
Db 227 SerSerLysThrThrAsnThrAsnSerProAsnSerIleSerLysThrSerIleLys 246
QY 621 -----ACCGTGACCATGCGGAATTTAGA----- 644
Db 247 GlnSerSerValAsnAlaSerGlyAsnValSerProSerGlnPheSerThrGlyAsnAsp 266
QY 645 AGGTGCTCTTGGCCACCTATGGCTGAGGTT----- 674
Db 267 AlaSerProThrSerProMetAlaSerLeuSerSerProLeuAsnThrAsnIleLeuGly 286
QY 675 -----CCTATGTTCTCTACTCAAAACTCCATGGCAGCTCGGCTTCTGCCTCC 722
Db 287 TyrProLeuSerProIleThrSerThrLeuGlyGlnAlaAsnThrSerThrSerThr 306
QY 723 GCTTCTTCATTAGAAATGTGGAAAGGATTGGAGGAGACTCAACTCTCTATCGATCAT 782
Db 307 AlaAlaThrThrLysThr---AspAlaAspThrProSerThrMetAsnThrAsnAsnAsn 325
QY 783 GACATGAACACAAACAAATTTGGTTCGGCGAATAAATCATATGTCAACACGGGTAAG 842
Db 326 AsnAsnAsnAsnAsn-----SerAlaAsnLeuAsnAsnIle-----ProGlnArg 340
QY 843 GTCGAGGAATGGACTTCTAAGTTCCTTTCATCACTCTTTTCTTTCTCTTCTCCATTTCC 902
Db 341 IlePheSerLeuAspAspIleSerSerPheAsnSerSerArgLysSerLeuAsnLeuAsp 360
QY 903 CACTAGTTCTGTTCTTCTCTCTAGATACCCCTCTTTTCAGGGGACTCTCGTCCCTACT 962
Db 361 Asp---SerAsnSerLeuPheLeuTrpAspThrSerGlnHisSerAsnAlaSerMetThr 379
QY 962 ----- 962
Db 380 AsnThrAsnMetHisAlaGlyValAsnAsnSerGlnSerGlnAsnAspGlnSerSerLeu 399
QY 963 -----ATTGTTGTCATTCTCGAAACATCTCTCCCGTGCATTTTCTCTTCCCTTATAT 1016
Db 400 AsnTyrMetGluAsnIleMetGluLeuTyrSerAsn-----Tyr 412
QY 1017 ACATATATATATATATATATATATATCTCTCTCTCTAGGTATTTTGTATTTCTGTGCTT 1076
Db 413 ThrGlySerGluLeuSerSerHisThrAlaIleLeuArgPheLeuValValLeuThrLeu 432

RESULT 10

US-08-724-974A-3
; Sequence 3, Application US/08724974A
; Patent No. 5912335
; GENERAL INFORMATION:
; APPLICANT: Derk J. Bergsma, Catherine E. Ellis
; TITLE OF INVENTION: A NO. 5912335el G-Protein Coupled Receptor
; TITLE OF INVENTION: HUVCT36
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road, P.O. Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:

Db 271 AsnSerGlnThrLysArgTyrSerGlyAsnMetGluTyrValIleSerArgIleSerPro 290
QY 405 GTTTCGCCGATTCGCGCTCAGCGGTATAAAAGAGATCTTTTTTCTCGGCTGCC 464
Db 291 ValProProGlyAlaIleProGlnGluGlyTyrProPro----- 302
QY 465 CTTCCATTTTAAAT-----GTCCTTA 485
Db 303 ---ProProLeuAsnThrSerProMetAsnProProAsnGlnGlyGlnArgGlyIleSer 321
QY 486 TCTGCTCTCTTGATCTTACGGTCTCACTACCTCTCTTCAACGCTCAATAATTTCCC 545
Db 322 SerValProValGlyArgGlnProIleMetGlnSerSerSerLysPheAsnPhePro 341
QY 546 GCT-----ATCAAAAT-----TCCCAAGACTACTTTTACGCTCAAAAT 584
Db 342 SerGlyArgProGlyMetGlnAsnGlyThrGlyGlnThrAspPheMetIleHisGlnAsn 361
QY 585 -----CGCTGCCAA 593
Db 362 ValValProAlaGlyThrValAsnArgGlnProProProProTyrProLeuThrAlaAla 381
QY 594 CAACAACAGCCCTTCCACATTCGCTACC-----GTGACCATCGCGAATTTAGA 644
Db 382 AsnGlyGlnSerProSerAlaLeuGlnThrGlyGlySerAlaAlaProSerSerTyrThr 401
QY 645 AGGGTGCTTTGCCACCTATG----- 665
Db 402 AsnGlySerIleProGlnSerMetMetValProAsnArgAsnSerHisAsnMetGluLeu 421
QY 666 -----GTGAGGTTCCTATGTTGTACTCAAACTCCATCGGCGAGCTCGCTTCTGCC 719
Db 422 TyrAsnIleSerValProGlyLeuGlnThrAsnTpproGlnSerSerSerAlaProAla 441
QY 720 TCCGCTTCT-----TCATTAGAAATGTGGGAAAGCATTTGGAGAGAGA 764
Db 442 GlnSerSerProSerSerGlyHisGluIleProThrTrpGlnProAsnIleProValArg 461
QY 765 CTCAACTCTATCGATCATGACATGAACAAC 794
Db 462 SerAsnSerPheAsnAsnProLeuGlyAsn 471

RESULT 12

US-09-134-001C-5595
; Sequence 5595, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5595
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5595
Alignment Scores:
Pred. No.: 0.442 Length: 151
Score: 78.00 Matches: 37
Percent Similarity: 41.22% Conservative: 24
Best Local Similarity: 25.00% Mismatches: 41
Query Match: 3.77% Indels: 46
DB: 4 Gaps: 7
US-09-814-661A-1 (1-1158) x US-09-134-001C-5595 (1-151)

QY 707 CTCGCTTCTGCCGCTTCTTCAATAGAAATGTGGAAAGAGATTTTGGAGGACAGACT 766
Db 24 LeuLysPhe-----LeuHisPhePheLeuIleHisValIleLysSerIleAsp----- 39
QY 767 CAACTCTATCGATCATGACATGAACAACAATAATTTGGTTTCTGGCGAACTAAAAATCAT 826
Db 40 -----HisPheThrCysThrPheValTyrTyrSerThrLysPheTyr 53
QY 827 GTTCACCCAGGTAGGTCGAGGAANTGAGCTTCTAAAGTTCTCTTCACTACTCTTTCTT 886
Db 54 -----ArgSerSerGlyCysIleSerTyrLysValPheAsnIleIle 67
QY 887 TTTCTTTTCCATTTCCCACTAGTTCTGTTCTTTCTTCTCTTAGATACCTTCTTTTTCAG 946
Db 68 Phe-IleSerIleAlaSerIleTyrIleCysPheLysPheLysAsnLeuLeuIle-- 86
QY 947 GGACTCTCGCTCTACTATTGTTCATTTCTGAAACATTTCTCCCGTCC----- 996
Db 87 -----AsnProPheIleAsnValIleVal-ArgAspAsnLeuPheCysPheLeuSerG 104
QY 997 -----ATTTTCTCTTCCCTTTAT----- 1014
Db 104 IyAsnLeuPheLeuSerCysPheLeuPheLeuIleThrIleArgIleCysTyrAlaProI 124
QY 1015 -----ATACATATA-----TATATATATATATATATATATCTCTTCT 1051
Db 124 IeSerLeuHisIleIleArgPheIleLeuPhePheTrpIleAlaPheCysPhePheT 144
QY 1052 ACGTATTTTGTATTCTGT 1071
Db 144 yrIlePheLeuSerThrCys 150
RESULT 13
US-08-833-488B-24
; Sequence 24, Application US/08833488B
; Patent No. 6060326
; GENERAL INFORMATION:
; APPLICANT: Frank, Glenn R.
; APPLICANT: Rushlow, Keith E.
; TITLE OF INVENTION: Method to Detect Canine Ige
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,488B
; FILING DATE: 07-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: DI-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-833-488B-24


```
QY 744 GAAAAGGATTTCGAGGAGAGACTCACTCTATCGATCATGACATGAACAACAACAATTT 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 -IleArgTyrTyrGluAsnPheAsnIleSerIleSerAsnValThrThrLysAsn-- 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 804 GGTTCGGCACTAAATCTATGTTCAACACGAGGTAGGTGCGAGAAATGGACTTCTAA 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 -----SerGlyAsnTyrSerCysSerGlyGlnIleGlnLysGlyTyrThrSerly 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 864 AGTTCCTTTTCATACACTC-----TTTTC 884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 sValLeuAsnIleIleValLysLysGluProThrLysGlnAsnLysTyrSerGlyLeuGl 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 885 TTTTCTCTTCCATTTCACACTAGTCTCTCTCTTCTCTCTTAGATACCCCTCTTTTC 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 nPheLeuIleProLeu--ValValIleLeuPhe--AlaValAspThrGlyLeuPhe 226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 945 AGGACTCTCGCTCTACTATTGTTGTCATTCTCGAA 980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227 IleSerThrLysGlnGlnLeuThrValLeuLeuGln 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-134-001C-4149
; Sequence 4149, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4149
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4149

Alignment Scores:
Pred. No.: 1.98 Length: 418
Score: 75.00 Matches: 42
Percent Similarity: 40.70% Conservative: 28
Best Local Similarity: 24.42% Mismatches: 76
Query Match: 3.68% Indels: 26
DB: 4 Gaps: 7

US-09-814-661A-1 (1-1158) x US-09-134-001C-4149 (1-418)
QY 1094 TATAGACTATCTTTGATAAGACACAGAATAACAAAATACGTAGAGACATATATAT 1035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TyrAsnValThrGlnHisAlaThrTyrLysThrLysAsn-----LysArgGluThrAla 21
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1034 ATATATATATATATATATATAAGGAAAGGAAATGCACGGGAGAGAAATGTTTCGAG 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22 ValLeuIleGlyValHisAlaGlnThrAspArgGlnPheAsnPheGluSerThrMetGlu 41
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 974 AAT---GACAACAATAGTAGGACGAGAGTCCCTGAAAAGAGGGGTATCTTAGAGAGAGAAA 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42 GluLeuAlaLeuSerGlnThrCysGlnLeuAsnValLysGlyGlnIleThrGlnAsn 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 917 AGACAGAACTA-----GTGGGAATCGAAAGACAGAGAAAAGAGACTATG 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 ArgGluGlnPheAspHisLysTyrTyrValGlyLysGlyLysIleAspGlu-----Ile 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 872 AAAGGAACCTTAGAAGCTCATTTCTCTCGACCTTACCCCTGTTGCAACATAGATTTTAGTTC 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 LysSerPheIleGluPheHisAspIleAspValValIleThrAsnAspGluLeuThrThr 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 GCCAGAACCAAAATTTCTTGTTCATGTC-----ATGATCGATAGAGTTGAGTCT 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 100 AlaGlnSerLysThrLeuAsnAspAsnLeuGlyIleLysIleLysArgThrGlnLeu 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 761 CTCCTCCAAATCCTTTTCCACATTTCTAATGAAGAAGCGGAGGAGCGAGCTGCC 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 IleLeuGluIlePhe-----AlaLeuArgAlaArgSerArg 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 701 CATGGAGTTTTCAGTAGACAACATAGGAACCTCAGCCATAGCTGGCAAGGACCCCTTCT 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 GluGlyLysLeuGlnValGluLeuAlaGlnLeuAspTyrLeuLeuProArg----- 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 641 AAATTCGCCCATGGTCACGGTACGCAATGTGGAAGG 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 -----LeuHisGlyHisGlyLysSerLeuSerArg 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 16, 2003, 12:03:17
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 17:04:17 ; Search time 211 Seconds
(without alignments)
1109.990 Million cell updates/sec

Title: US-09-814-661a-2

Perfect score: 537

Sequence: 1 MNSQDYFYAQRNCRQQQAP.....FGSGELKSMFNCKGKVEEMDF 104

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O/Cgn2.1/USPTO.spool/US09814661/runat_16072003.115309.14262/app_query.fasta_1.263
-DB=N_Genseq_101002 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=p2n -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US09814661 -ECGN_1.1.396 -runat_16072003.115309.14262 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Genseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	537	100.0	1158	21	AAAL4263
2	82	15.3	5113	23	ABL14387
3	82	15.3	8700	23	ABL14386
4	80	14.9	5286	23	AS73156
5	80	14.9	5954	23	AS80591
6	80	14.9	6143	23	AS83843
7	76.5	14.2	576	24	ABO90913
8	75	14.0	1260	23	ABL16225
9	75	14.0	1683	22	AAH77996
10	75	14.0	4978	23	ABL16224
11	75	14.0	7544	23	ABL14763
12	74.5	13.9	20348	23	ABL14606
13	73.5	13.7	1278	22	AAH66089
14	73.5	13.7	34980	22	AAH68527
15	73	13.6	3409	23	ABL11291
16	73	13.6	5467	23	ABL11290
17	72	13.4	1116	23	AS77928
18	72	13.4	2403	23	AA577931
19	72	13.4	8931	21	AA575307
20	71.5	13.3	987	23	ABL24801
21	71.5	13.3	1233	22	AAH66114
22	71.5	13.3	3089	23	ABL24800
23	71.5	13.3	34980	22	AAH68527
24	71	13.2	859	21	AAF07571
25	71	13.2	2436	24	ABL91361
26	71	13.2	5502	23	ABL28649
27	71	13.2	7715	23	ABL21006
28	71	13.2	8454	23	ABL24012
29	71	13.2	10234	23	ABL28648
30	71	13.2	1230025	20	AA591990
31	70.5	13.1	498	22	AAH19125
32	70.5	13.1	498	22	AAH19144
33	70.5	13.1	538	24	ABO53536
34	70.5	13.1	538	24	ABO53537
35	70.5	13.1	611	13	AAQ27224
36	70.5	13.1	2249	22	AAK89395
37	70.5	13.1	2250	22	AAK89396
38	70.5	13.1	2250	22	AAK89397
39	70.5	13.1	2989	23	ABL25392
40	70.5	13.1	5381	22	AAF30037
41	70.5	13.1	18660	21	AA58472
42	70.5	13.1	81001	22	AAF30035
43	70.5	13.1	4403765	22	AAI99683
44	70.5	13.1	4411529	22	AAI99682
45	70	13.0	531	23	ABL11263

ALIGNMENTS

RESULT 1

AAAL4263

ID AAAA14263 standard; DNA: 1158 BP.

XX

AC AAAA14263;

XX

DT 15-AUG-2000 (first entry)

XX Yeast SML1 (suppressor of mecl lethality) gene.

DE SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;

KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;

KW deoxynucleotide triphosphate synthesis; Mecl; Rad53;

KW cell cycle checkpoint; cancer; microbial infection;

KW ataxia telangiectasia; AT; ds.

XX Saccharomyces cerevisiae.

OS

XX

FH Key Location/Qualifiers
 FT CDS 549..863
 FT /tag=a
 FT /product="Yeast Sm11 protein"
 FT /function="Inhibitor of ribonucleotide reductase"
 XX
 XX WO200017225-A2.
 XX
 XX 30-MAR-2000.
 XX
 XX 24-SEP-1999; 99WO-US22260.
 XX
 XX 24-SEP-1999; 98US-0158858.
 XX
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 XX Rothstein R, Zhao X;
 XX WPI: 2000-283539/24.
 XX P-PSDB; AAY90470.
 XX
 XX New Sm11 protein and its homologs, useful for treating cancer, for
 XX microbial infection and ataxia telangiectasia and in screening for
 XX specific modulators
 XX
 XX Claim 6; Fig 1C; 98pp; English.
 XX
 XX This sequence represents the yeast SML1 (suppressor of mecl lethality)
 XX gene. This gene, located on chromosome XIII, encodes Sm11, an inhibitor
 XX of ribonucleotide reductase. The SML1 gene (assigned the name YML058w by
 XX the Stanford yeast genomic project) was initially isolated in a mecl-1
 XX strain by genetic analysis. Sm11 is a suppressor of the mecl mutant in
 XX yeast which is associated with abnormal levels of recombination in both
 XX meiosis and mitosis. Sm11 also permits cell growth in the absence of the
 XX cell cycle checkpoint proteins Mec1 and Rad53, which, unlike most
 XX checkpoint proteins, are essential for cell growth. Sm11 binds to
 XX the largest subunit of ribonucleotide reductase (RNR) which catalyses the
 XX rate-limiting step of deoxynucleotide triphosphate (dNTP) synthesis,
 XX conversion of NTPs to dNTPs. Binding of Sm11 reduces the activity of RNR
 XX and thus inhibits the synthesis of dNTPs. The SML1 gene, or fragments
 XX thereof may be used to isolate the human homologue of Sm11. The Sm11
 XX protein, and compounds that modulate the interaction of Sm11 with
 XX ribonucleotide reductase (RNR), may be used to alter the rate at which
 XX cells divide. These are particularly useful for treating cancer,
 XX microbial infection and ataxia telangiectasia (AT), a condition in which
 XX patients are prone to cancer. Sm11 may also be used to screen for
 XX modulatory agents, to raise specific antibodies, and for stimulating the
 XX function of the ATM gene (a mammalian Mec1 homologue which is mutated in
 XX AT). Anti-Sm11 antibodies are used as diagnostic and analytical
 XX immunoassay reagents and to remove Sm11 from serum or to titrate Sm11
 XX intracellularly.
 XX
 XX SQ Sequence 1158 BP; 289 A; 282 C; 196 G; 391 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,97e-56 Length: 1158
 Score: 537.00 Matches: 104
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-814-661A-2 (1-104) x AAA14263 (1-1158)
 QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 Db 549 ATGCAAAATTCGCAAGACTACTTTTACGCTCAAAATTCGCTGCAACAAACAGCCCT 608
 QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuProMetAla 40
 Db 609 TCCAAATTCGGTACCGTACCAATGCGGGAATTTAGAAAGGTGCTTTGCCACCTATGGCT 668
 QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
 Db 669 GAGGTTCTATGTGTCTACTCAAACTCCATGGCAGCTCGCTTCTCGCTCCGCTTCT 728
 QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisaspMet 80
 Db 729 TCATTAGAAATGTGGGAAAGGATTTGGAGGAGAGACTCAACTCTTATCGATCATGACATG 788
 QY 81 AsnAsnAsnLysPheGlySerGlyCyluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
 Db 789 AACAAACAACAATTTGGTTCTGGCGAAGCTAAATCTATGTTCACACCGGTAAGGTCGAG 848
 QY 101 GluMetAspPhe 104
 Db 849 GAAATGGACTTC 860
 RESULT 2
 ABL14387
 ID ABL14387 standard; cDNA; 5113 BP.
 XX
 XX AC ABL14387;
 XX
 XX DT 26-MAR-2002 (first entry)
 XX
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 37643.
 XX
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical; gene; ss.
 XX
 XX OS Drosophila melanogaster.
 XX
 XX PN WO200171042-A2.
 XX
 XX PD 27-SEP-2001.
 XX
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX
 XX PR 23-MAR-2000; 2000US-191637P.
 XX
 XX PR 11-JUL-2000; 2000US-0614150.
 XX
 XX PR (PEKE) PE CORP NY.
 XX
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX DR WPI: 2001-656860/75.
 XX
 XX DR P-PSDB; ABB70284.
 XX
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions
 XX
 XX PS Claim 1; SEQ ID NO 37643; 21pp + Sequence Listing; English.
 XX
 XX CC The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX sequences (ABL01840-ABL16175) and the encoded proteins
 XX (ABB57737-ABB72072).
 XX
 XX CC The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 5113 BP; 1375 A; 1523 C; 1433 G; 782 T; 0 other;
 Alignment Scores:
 Pred. No.: 18.5 Length: 5113
 Score: 82.00 Matches: 28
 Percent Similarity: 43.88% Conservatives: 15
 Best Local Similarity: 28.57% Mismatches: 41
 Query Match: 15.27% Indels: 14
 DB: 23 Gaps: 3

Sequence 8700 BP: 1898 A: 2006 C: 2240 G: 2556 T: 0 other:

polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping and in recombinant production of (II). The

polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 5286 BP; 1786 A; 1178 C; 1073 G; 1249 T; 0 other;

Alignment Scores:

Pred. No.:	34.1	Length:	5286
Score:	80.00	Matches:	41
Percent Similarity:	44.63%	Conservative:	13
Best Local Similarity:	33.88%	Mismatches:	35
Query Match:	14.90%	Indels:	32
DB:	23	Gaps:	9

US-09-814-661A-2 (1-104) x AAS73156 (1-5286)

Qy 2 GlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaProSer 21
 ::::|||||:::|||||
 Db 4450 GAGAATTCCTAAAGACGTCACAGCAGCTCAGAAATTTAGTAAGAGAATCAGAGGCCCATCA 4509
 ::::|||||:::|||||
 Qy 22 -----ThrLeuArgThrValThrMetAlaGluPhe-----ArgArgValPro 35
 |||||
 Db 4510 CCCATCATTCCACAGCCTCAGGAGCAGAGAAATTCCTGACAATCAGAGAGGCTGAGC 4569
 :|||
 Qy 36 LeuProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAla 55
 |||||
 Db 4570 ---CCTCTTTT-----CCACTGGAGCCTGCACAGAAATCTAGAGTAAGCAGTCCA 4617
 :|||
 Qy 56 SerAla-----SerAlaSerSerLeuGluMetTrpGluLysAsp 68
 |||||
 Db 4618 CTGCCCAAGTTTCTGCAACAAAGAGGAGTCTTCATCTCTGGAG---TGGGAACCTGAG 4674
 :|||
 Qy 69 -----LeuGluGluArgLeuAsnSerIleAsp---HisAspMetAsnAsn 83
 :|||
 Db 4675 CCACACCTCTATCGTTTCNAAGAGTTTAAAGCATTATGTTCAT----- 4719
 :|||
 Qy 84 LysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGluGluMetAsp 103
 |||||
 Db 4720 -----GGCGATCTACTACGAAAGGCCATCTCTCAAAAGTCAGGAGCGGCAT 4767
 :|||
 Qy 104 Phe 104
 |||
 Db 4768 TTT 4770

RESULT 5
 AAS80591
 ID AAS80591 standard; cDNA; 5954 BP.
 XX AAS80591;
 AC AAS80591;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #16395.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS
 XX

PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB: ABG16404.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1; SEQ ID No 16395; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 5954 BP; 2007 A; 1302 C; 1219 G; 1426 T; 0 other;

Alignment Scores:

Pred. No.:	40.2	Length:	5954
Score:	80.00	Matches:	41
Percent Similarity:	44.63%	Conservative:	13
Best Local Similarity:	33.88%	Mismatches:	35
Query Match:	14.90%	Indels:	32
DB:	23	Gaps:	9

US-09-814-661A-2 (1-104) x AAS80591 (1-5954)

Qy 2 GlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaProSer 21
 ::::|||||:::|||||
 Db 5124 GAGAATTCCTAAAGACGTCACAGCAGCTCAGAAATTTAGTAAGAGAATCAGAGGCCCATCA 5183
 ::::|||||:::|||||
 Qy 22 -----ThrLeuArgThrValThrMetAlaGluPhe-----ArgArgValPro 35
 |||||
 Db 5184 CCCATCATTCCACAGCCTCAGGAGCAGAGAAATTCCTGACAATCAGAGAGGCTGAGC 5243
 :|||
 Qy 36 LeuProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAla 55
 |||||
 Db 5244 ---CCTCTTTT-----CCACTGGAGCCTGCACAGAAATCTAGAGTAAGCAGTCCA 5291
 :|||
 Qy 56 SerAla-----SerAlaSerSerLeuGluMetTrpGluLysAsp 68
 |||||
 Db 5292 CTGCCCAAGTTTCTGCAACAAAGAGGAGTCTTCATCTCTGGAG---TGGGAACCTGAG 5348
 :|||

QY 69 -----LeuGluGluArgLeuAsnSerIleAsp---HisAspMetAsnAsn 83
 Db 5349 CCACACCTCTATCGTTCAAGAGTTTAAAAAGCATTAAATGTTTCAT----- 5393
 QY 84 LysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGluGluMetAsp 103
 Db 5394 -----GGCGATCTACTACGAAAAAGCCATCTCCAAAGATCAGGGAGGCCCAT 5441
 QY 104 Phe 104
 Db 5442 TTT 5444
 RESULT 6
 ID AAS83843/c
 ID AAS83843 standard; cDNA; 6143 BP.
 AC AAS83843;
 XX 13-FEB-2002 (first entry)
 DT
 DE DNA encoding novel human diagnostic protein #19647.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG19656.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 19647; 103pp; English.
 CC
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6143 BP; 1489 A; 1251 C; 1342 G; 2061 T; 0 other;

Alignment Scores:
 Pred. No.: 42 Length: 6143
 Score: 80.00 Matches: 41
 Percent Similarity: 44.63% Conservatives: 13
 Best Local Similarity: 33.88% Mismatches: 35
 Query Match: 14.90% Indels: 32
 DB: 23 Gaps: 9
 US-09-814-661A-2 (1-104) x AAS83843 (1-6143)
 QY 2 GlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnGlnAlaProSer 21
 Db 1020 GAGAATTTCTAAAGACGCTCAGCAGCTCAGAATTTAGTAAGAGAAATCAGGAGCCCATCA 961
 QY 22 -----ThrLeuArgThrValThrMetAlaGluPhe-----ArgArgValPro 35
 Db 960 CCCATCACATTCACCGAGCTCAGGGAAGCAGAATTTCTGACAATCAGAGGAGGCTGAGC 901
 QY 36 LeuProProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAla 55
 Db 900 -----CTTCCTTT-----CCACTGGAGCCTGCACAGAAATCTAGAGTACGAGTCCA 853
 QY 56 SerAla-----SerAlaSerSerLeuGluMetTrpGluLysAsp 68
 Db 852 CTGGCCAGTTTCTGCAGCAACAAGGAGCTGCTTCTATCTCTGAG---TGGGAACCTGAG 796
 QY 69 -----LeuGluGluArgLeuAsnSerIleAsp---HisAspMetAsnAsn 83
 Db 795 CCACACCTCTATCGTTCAAGAGATTTAAAAAGCATTAAATGTTTCAT----- 751
 QY 84 LysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGluGluMetAsp 103
 Db 750 -----GGCGATCTACTACGAAAAAGCCATCTCCAAAGATCAGGGAGGCCCAT 703
 QY 104 Phe 104
 Db 702 TTT 700
 RESULT 7
 ABQ90913/c
 ID ABQ90913 standard; DNA; 576 BP.
 XX
 AC ABQ90913;
 XX
 DT 01-OCT-2002 (first entry)
 XX
 DE M. capsulatus gene #898 for DNA array.
 XX
 KW Micro array; gene; ds; differential expression; gene expression.
 OS Methylococcus capsulatus.
 XX
 PN WO200255655-A2.
 XX
 PD 18-JUL-2002.
 XX
 PF 14-JAN-2002; 2002WO-NO00019.
 XX
 PR 12-JAN-2001; 2001NO-0000235.
 PR 12-JAN-2001; 2001NO-0000239.
 XX
 PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
 PA (TIGR-) TIGR.
 XX
 PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
 PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
 PI Salzberg SL;
 XX
 DR WPI; 2002-557818/59.
 XX
 PT Novel DNA array useful for determining differential expression of
 PT Methylococcus capsulatus genes, comprises polynucleotides or

PT oligonucleotides representative for a selective number of *Methylococcus*
XX capsulatus genes
PS
XX
XX Claim 14; Page 388; 678pp; English.
XX
CC The invention relates to a novel DNA array giving a representation of a
CC number of *Methylococcus capsulatus* genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC *M. capsulatus*, and for studying gene expression on a genomic scale and in
CC gene expression assays of *M. capsulatus* genes. The sequences shown in
CC ABQ90016-ABQ91855 represent *M. capsulatus* genes for use in arrays of the
CC invention.
XX
SQ Sequence 576 BP; 119 A; 183 C; 168 G; 106 T; 0 other;

Alignment Scores:
Pred. No.: 4.32 Length: 576
Score: 76.50 Matches: 21
Percent Similarity: 46.77% Conservative: 8
Best Local Similarity: 33.87% Mismatches: 26
Query Match: 14.25% Indels: 7
DB: 24 Gaps: 1

US-09-814-661A-2 (1-104) x ABQ90913 (1-576)

QY 8 PheTyAlaGlnAsnArgCysGlnGlnGln-----AlaPro 20
DB 218 TTCATGCTCCGGGAACATGCCGCCACCATTCGCCCATGCGCGGATGAGGGCCGCCCG 159

QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
DB 158 GCGAGCGTTTCAGGTTTCAGCGTGAATTCACCGAGAAGATCGCCCTGGGCCCATAGGCC 99

QY 41 GluValProMetLeuSerThrClnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
DB 98 ACGGACCGATGCTGAGCATCGATGATGCGGGCGGGGTATCGGGCCGCTCGCTTCGATGCT 39

QY 61 SerLeu 62
DB 38 ACGCTG 33

RESULT 8
ABL16225
ID ABL16225 standard; DNA; 1260 BP.
XX
AC ABL16225;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 148.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
PN
PN WO200171042-A2.
PD
PD 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell

PT interactions -
XX
PS Claim 1; SEQ ID NO 148; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL161175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1260 BP; 338 A; 373 C; 309 G; 240 T; 0 other;

Alignment Scores:
Pred. No.: 19.5 Length: 1260
Score: 75.00 Matches: 23
Percent Similarity: 51.65% Conservative: 24
Best Local Similarity: 25.27% Mismatches: 40
Query Match: 13.97% Indels: 4
DB: 23 Gaps: 1

US-09-814-661A-2 (1-104) x ABL16225 (1-1260)

QY 17 GlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeu 36
DB 551 AGCGGAGTCAGTCACATTCGCCAGCAGCAGTCCAGAGCAGTCATATGGAAGCGACT 610

QY 37 Pro-MetAlaGluVal-ProMetLeuSerThrGlnAsnSerMetGlySerSerAla 56
DB 611 CCTCCGAGGAGCAGTCGCCCGCAGTCAGTAGCCACCATCTGTGTTTAAACGCCGCTG 670

QY 56 eAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSer 76
DB 671 CAGCGCTGCTCGCGCGGTCCACCTCAAGTCCACGGCGATCAGAGTAACCTCTCGCCAA 730

QY 76 LeAspHisAspMetAsnAsn-----LysPheGlySerGlyGluLeuLysSerMet 94
DB 731 TAGCGATTCAGGTGCAAAATAACCTTCGCAACTACGGACAGGATCCTTAAACGCCCTTAT 790

QY 94 heAsnGlnGlyLysValGluGluMetAsp 103
DB 791 GTGGATCAAGCCCAAGCAGGAGATGGAT 819

RESULT 9
AAH77996
ID AAH77996 standard; DNA; 1683 BP.
XX
AC AAH77996;
XX
DT 13-NOV-2001 (first entry)
XX
DE Nucleotide sequence of human protein kinase SGK424.
XX
KW Human; protein kinase; cancer; immune disease; cardiovascular disease;
KW brain disease; neuronal disease; Alzheimer's disease; chromosome 19;
KW Parkinson's disease; multiple sclerosis; metabolic disorder;
KW peripheral nervous system disease; amyotrophic lateral sclerosis;
KW infection; ocular disease; migraine; pain; sexual dysfunction;
KW mood disorder; attention disorder; cognition disorder; hypotension;
KW hypertension; psychotic disorder; dyskinesia; transplant rejection; ss.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT CDS 1..1683
XX FT /tag= a
XX FT /product= "protein kinase"

PN WO200166594-A2.
XX 13-SEP-2001.
PD
DE
XX 02-MAR-2001; 2001WO-US06838.
PF
KW 06-MAR-2000; 2000US-0187150.
XX 29-MAR-2000; 2000US-0193404.
PR 13-NOV-2000; 2000US-0247013.
XX (SUGEN-) SUGEN INC.
PA
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI
XX WPI; 2001-536777/59.
DR P-PSDB; AAG67397.
XX
XX Nucleic acids capable of encoding human polypeptides having a kinase or
PT kinase-like activity, useful for diagnosing a disease selected from
PT cancers, cardiovascular disease and neuronal-associated diseases (e.g.
PT Alzheimer's disease) -
XX
XX Example 1; Fig 1G-H; 201pp; English.
PS
XX The present sequence encodes a human protein kinase. The gene is
CC located at chromosomal position 19q12-q13.3. The kinase polypeptides
CC are useful for diagnosing a disease or disorder selected from cancers
CC (e.g. cancers of tissues and cancers of hematopoietic origin),
CC immune-related diseases and disorders, cardiovascular disease, brain
CC or neuronal-associated diseases (e.g. Alzheimer's disease, Parkinson's
CC disease, multiple sclerosis), metabolic disorders, peripheral nervous
CC system diseases, amyotrophic lateral sclerosis, viral infections,
CC infections caused by prions, infections caused by bacteria, infections
CC caused by fungi, ocular diseases, migraines, pain, sexual dysfunction,
CC mood disorders, attention disorders, cognition disorders, hypotension,
CC hypertension, psychotic disorders, dyskinesias, and organ transplant
CC rejection. Kinase inhibitors are useful for treating diseases and
CC disorders described above.
XX
XX Sequence 1683 BP; 292 A; 566 C; 528 G; 297 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 29 Length: 1683
Score: 75.00 Matches: 23
Percent Similarity: 46.88% Conservative: 7
Best Local Similarity: 35.94% Mismatches: 32
Query Match: 13.97% Indels: 2
DB: 22 Gaps: 1
US-09-814-661A-2 (1-104) x AAH77996 (1-1683)
QY 36 LeuProProMetAlaGluValPromMetLeuSerThrGlnAsnSerMetGlySerSerAla 55
Db TTGCCACCCATGCGATCTCTGGACCATGTG-----CAGCCAGATGTTGGACACCCG 621
QY 56 SerAlaSerAlaSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSer 75
Db CTGGCTCCCTGCTCCCTGAGGTGATGATGAGGCAGAGCAGTGTGAGGTAGCACAG 681
QY 76 IleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsn 95
Db ATCAGAGAGACTCCCAAGTCCCGGCTTGGGGTTTGGTCACTGAGTAGCCTGTGGCCA 741
QY 96 GlnGlyLysVal 99
Db CCGGGGCTGGTA 753
Db 742 CCGGGGCTGGTA 753
RESULT 10
ABL16224
ID ABL16224 standard; DNA; 4978 BP.
XX ABL16224;
AC ABL16224;
XX

DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 145.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 145; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4978 BP; 1574 A; 1041 C; 926 G; 1437 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 130 Length: 4978
Score: 75.00 Matches: 23
Percent Similarity: 51.65% Conservative: 24
Best Local Similarity: 25.27% Mismatches: 40
Query Match: 13.97% Indels: 4
DB: 23 Gaps: 1
US-09-814-661A-2 (1-104) x ABL16224 (1-4978)
QY 17 GlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeu 36
Db AGGCGAGTCCAGTACATTCGCGACGACGAGTGCATATGGAAGCGACT 3328
QY 37 Pro-PrometAlaGluVal-PrometLeuSerThrGlnAsnSerMetGlySerSerAla 56
Db CCTCCGAGAACGCGATCGCCCGAGTCAGTACACCTGTGTTTAAACGCCGTG 3388
QY 56 eAlaSerAlaSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerI 76
Db CAGCCGCTGCTGGCGGTCTCCACCTCAAGTCCACGGCGATGAGAGATAACCTCTCCCAA 3448
QY 76 leAspHisAspMetAsnAsn-----LysPheGlySerGlyGluLeuLysSerMetP 94
Db TAGGCGATCAGGTGCAAAATAACCTTCGCAACTACGCAACAGATCCTTAAACGCCCTTAT 3508
QY 94 heAsnGlnGlyLysValGluGluMetAsp 103
Db GTGGGATCAAGCCCAAGCAGGAGATGGAT 3537

```
RESULT 11
ABL14763
ID ABL14763 standard; cDNA; 7544 BP.
XX
AC ABL14763;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 38771.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB70660.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 38771; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7544 BP; 1870 A; 2160 C; 1911 G; 1603 T; 0 other;

Alignment Scores:
Pred. No.: 231 Length: 7544
Score: 75.00 Matches: 26
Percent Similarity: 59.4% Conservative: 21
Best Local Similarity: 32.91% Mismatches: 28
Query Match: 13.97% Indels: 4
DB: 23 Gaps: 2

US-09-814-661A-2 (1-104) x ABL14763 (1-7544)

QY 9 TyrAlaGlnAsnArgCysGlnGlnGlnGlnAlaProSerThrLeuArgThrVal-ThrMe 28
Db 2689 TGGGTCGCATCACACAGGTCGCAAGAGAGTCGATGGTGGTGGAGAACATGAAACCG 2748

QY 28 talaGluPheArgValProLeuPro-PrometAlaGluValProMetLeuSerThrG 48
Db 2749 TGACGAATTCAGCGACCTCCGCTCCGATATTTCAGGAGCTGCCAATGTTTCGGCGC 2808

QY 48 InAsnSerMetGlySerSerAlaSerAlaSerAlaSerSerLeuGluMet---TrpGluL 67
Db 2809 AA---TCCTTTGGGGCCAGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2865

RESULT 12
ABL14606
ID ABL14606 standard; cDNA; 20348 BP.
XX
AC ABL14606;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 38300.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB70503.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 38300; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 20348 BP; 5898 A; 4506 C; 4345 G; 5599 T; 0 other;

Alignment Scores:
Pred. No.: 1.05e+03 Length: 20348
Score: 74.50 Matches: 23
Percent Similarity: 50.00% Conservative: 13
Best Local Similarity: 31.94% Mismatches: 31
Query Match: 13.87% Indels: 5
DB: 23 Gaps: 2

US-09-814-661A-2 (1-104) x ABL14606 (1-20348)

QY 32 ArgArgValProLeuProMetAlaGluValProMetLeuSerThrGlnAsnSerMet 51
Db 5081 CGTCGCTTTCCGCCACCTCTCAGCGATTTTGTGCGCCACATCAGCGCAATCAACGACTG 5140

QY 52 GlySerSerAlaSerAlaSer-----AlaSerSerLeuGluMetTrpGlnLysAsp 68
Db 5141 CGGAGCGGAGCGCATATTTCTTTAATCGAGCCGCACTAATGACATGTTCCCTCGTCTC 5200
```


QY 69 LeuGluGluArgLeuAsnSerIleAspHisAspMetAsnAsnLysPheGlySerGly 88
 ...
 Db 5201 GTATGCAATCGACTCACTGACTCGACTTGAATCGAATCGTCTTTCCGACGG 5260
 QY 89 GluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
 ...
 Db 5261 CAAATTAATCGATT-----CAGGGGAGGATCGAG 5290

RESULT 13
 AAH66089/c
 ID AAH66089 standard; DNA; 1278 BP.
 XX
 AC AAH66089;

DT 26-SEP-2001 (first entry)
 DE C glutamicum coding sequence fragment SEQ ID NO: 1124.
 DE
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR P-PSDB; AAG90870.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS
 XX Claim 8; SEQ ID NO: 1124; 246pp + Sequence Listing; English.
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX

SQ Sequence 1278 BP; 295 A; 355 C; 348 G; 280 T; 0 other;
 Alignment Scores:
 Pred. No.: 30.4 Length: 1278
 Score: 73.50 Matches: 20
 Percent Similarity: 46.67% Conservative: 8
 Best Local Similarity: 33.33% Mismatches: 17
 Query Match: 13.69% Indels: 15
 DB: 22 Gaps: 3

US-09-814-661A-2 (1-104) x AAH66089 (1-1278)
 QY 45 LeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSer----- 58
 ...
 Db 1054 TTATCGAGTCGGGTTCGCCGAGGCAAAAGCGATCGATCCATCGCGGTGATTC 995
 QY 59 ---AlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAsp 77
 ...
 Db 994 CTTTCATCAGATCTTCGCCGGTGGTCACAGGACGGCAA-----TCAACAGAT 947
 QY 78 HisAspMetAsnAsnLys-----PheGlySerGlyLeuLysSerMet 93
 ...
 Db 946 CACGATCAGGACAGCAAAAGATGACGCCATCATCTTTGGTCCAGGTGCGCCACTTTG 887

RESULT 14
 AAH68527/c
 ID AAH68527 standard; DNA; 349980 BP.
 XX
 AC AAH68527;

DT 26-SEP-2001 (first entry)
 DE C glutamicum coding sequence fragment SEQ ID NO: 7062.
 DE
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS
 XX Disclosure; SEQ ID NO: 7062; 246pp + Sequence Listing; English.
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX

SQ Sequence 349980 BP; 79725 A; 90426 C; 98918 G; 80911 T; 0 other;
 Alignment Scores:
 Pred. No.: 7.01e+04 Length: 349980
 Score: 73.50 Matches: 20
 Percent Similarity: 46.67% Conservative: 8

```
Best Local Similarity: 33.33% Mismatches: 17
Query Match: 13.69% Indels: 15
DB: 22 Gaps: 3

US-09-814-661A-2 (1-104) x AAH68527 (1-349980)

Qy 45 LeuSerThrGlnAsnSerMetGlySerAlaSerAlaSer----- 58
    ||||| ||| ||| ||||| |||
Db 161922 TTATCGACGTGGGTTCGCGAGGCCAAGCGCATCGCGATCCTCACCATCGCGGTGATTC 161863

Qy 59 ---AlaSerSerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAsp 77
    ||||| ||| ||| ||||| |||
Db 161862 COTTTCATCATGCTCTCGCGGTGGTTCACAGGACGCGCAA-----TCACACAGAT 161815

Qy 78 HisAspMetAsnAsnLys-----PheGlySerGlyGluLeuLysSerMet 93
    ||||| ||| ||| ||||| |||
Db 161814 CACGATCAGGACCAAGATGACGCCATCTTTGTCAGGTCCAGGTGCCCACTTTG 161755

RESULT 15
ABL11291
ID ABL11291 standard; cDNA; 3409 BP.
XX
AC ABL11291;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SFQ ID NO 28355.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI: 2001-656860/75.
DR P-PSDB; ABB67188.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 28355; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3409 BP; 1129 A; 701 C; 777 G; 802 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 136 Length: 3409
Score: 73.00 Matches: 29
Percent Similarity: 38.17% Conservative: 21
```

```
Best Local Similarity: 22.14% Mismatches: 35
Query Match: 13.59% Indels: 47
DB: 23 Gaps: 3

US-09-814-661A-2 (1-104) x ABL11291 (1-3409)

Qy 13 ArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArg 32
    ||||| ||||| ||||| ||||| |||||
Db 1210 AAATGTCAAAAANA-----TCGCTACGGCTTG-TCGGTTCTCGAGTTTCCC 1256

Qy 33 ArgValProLeuPro----- 37
    ||||| ||||| ||||| ||||| |||||
Db 1257 AAGGAGCCCAATGCCCAATCAATTTTGGGATTTTCGAAAAGTACAACCTTTGATTAT 1316

Qy 38 -----ProMetAlaGluValProMetLeuSerThrGlnAsnSerMet 51
    ||||| ||||| ||||| ||||| |||||
Db 1317 TACATTAAGGAGAGACAAATCCGCTACAGGAATGCCCTATGCTTAAAGAAATAGAAGT 1376

Qy 52 GlySerSerAlaSerAlaSerSerLeuGluMetTrpGluLysAsp----- 68
    ||||| ||||| ||||| ||||| |||||
Db 1377 AAAGAAATGTTGATGCTTCTTCAGTTGACCTTGGAATCTTCAAAAGCATCAGGATGCC 1436

Qy 69 -----LeuGlu 70
    ||||| ||||| ||||| ||||| |||||
Db 1437 AAAAGGAGAGATCCAGACAAACTACCAGTGGCCTTAGAAAAATGTTGAAACCCAAAAGAG 1496

Qy 71 GluArgLeuAsnSerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeu 90
    ||||| ||||| ||||| ||||| |||||
Db 1497 TTGGAGCTCGAAAACCTTACCAGAAAGTTACACATAATAATAAGATCTTTAGGGTCAAGAG 1556

Qy 91 LysSerMetPheAsnGlnGlyLysValGluGlu 101
    ||||| ||||| ||||| ||||| |||||
Db 1557 AAAGAAATGTTAAGGCAAGGAAGCTTAAAGAA 1589
```

Search completed: July 17, 2003, 18:15:04
Job time : 267 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:50:17 ; Search time 19 Seconds
(without alignments)
526.209 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 104
Sequence: 1 MONSODYFYAQNRCQQQAP.....FGSGELKSMFNQGVEMDF 104

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	104	2 S49803	hypothetical prote
2	9	8.7	2186	2 T13169	tiggrin - fruit fl
3	8	7.7	97	2 H86426	10.6k hypothetical
4	8	7.7	116	2 D96530	probable ribosomal
5	8	7.7	288	2 T05954	transcription fact
6	8	7.7	317	2 T00500	probable elicitor
7	8	7.7	368	2 T01597	hypothetical prote
8	8	7.7	393	2 G70929	probable pPE prote
9	8	7.7	474	2 I38240	transcription fact
10	8	7.7	514	2 D86280	protein T9E21.7 (i
11	8	7.7	522	2 B71807	DNA repair protein
12	8	7.7	524	2 A64694	DNA repair protein
13	8	7.7	538	2 E96492	hypothetical prote
14	8	7.7	591	2 AC3528	extracellular seri
15	8	7.7	670	1 S05091	myb-related protei
16	8	7.7	786	2 G85073	probable myosin-11
17	8	7.7	1111	2 T28972	hypothetical prote
18	7	6.7	66	2 A95115	hypothetical prote
19	7	6.7	66	2 D97984	hypothetical prote
20	7	6.7	124	2 A72629	hypothetical prote
21	7	6.7	133	2 T15449	hypothetical prote
22	7	6.7	150	2 A96647	hypothetical prote
23	7	6.7	159	2 T31598	hypothetical prote
24	7	6.7	174	2 T02683	hypothetical prote
25	7	6.7	192	2 H96828	hypothetical prote
26	7	6.7	203	2 S48995	hypothetical prote
27	7	6.7	217	2 B81067	conserved hypothet
28	7	6.7	217	2 G81801	probable lipoprote
29	7	6.7	229	2 E83548	ribonuclease III P

30	7	6.7	255	2 B97087	ABC-type transport
31	7	6.7	267	2 T45960	hypothetical prote
32	7	6.7	294	2 T48026	hypothetical prote
33	7	6.7	332	2 B91039	probable cytochrom
34	7	6.7	332	2 D85883	probable cytochrom
35	7	6.7	332	2 AG0809	probable membrane
36	7	6.7	332	2 A65015	yefH protein - Esc
37	7	6.7	334	2 S53490	RNA-binding protei
38	7	6.7	342	2 T45886	RNA-binding protei
39	7	6.7	343	2 T40306	hypothetical prote
40	7	6.7	354	2 A96596	hypothetical prote
41	7	6.7	396	2 JC5022	UDP-galactose tran
42	7	6.7	399	2 T49186	hypothetical prote
43	7	6.7	402	2 T04348	endosperm specific
44	7	6.7	409	2 T24138	hypothetical prote
45	7	6.7	410	2 E75290	probable multidrug

ALIGNMENTS

RESULT 1

S49803
hypothetical protein YML058w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YM9958.04
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 19-Apr-2002
C:Accession: S49803
R:Devlin, K.; Churcher, C.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49800
A:Accession: S49803
A:Molecule type: DNA
A:Residues: 1-104 <DEV>
A:Cross-references: EMBL:246729; NID:g577134; PIDN:CAA86717.1; PID:g577138; GSPDB:GNO
C:Genetics:
A:Gene: SGD:SM1; MIPS:YML058w
A:Cross-references: SGD:S0004523
A:Map position: 13L
C:Superfamily: Saccharomyces hypothetical protein YML058w

Query Match 100.0%; Score 104; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.7e-98;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONSODYFYAQNRCQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTONSMGSSASASAS 60
|||||
Db 1 MONSODYFYAQNRCQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTONSMGSSASASAS 60

QY 61 SLEMEKDLLEERLNSIDHDMNNKFGSGELKSMFNQGVEMDF 104
|||||
Db 61 SLEMEKDLLEERLNSIDHDMNNKFGSGELKSMFNQGVEMDF 104

RESULT 2

T13169
tiggrin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
R:Fogerty, F.J.; Fessler, L.I.; Bunch, T.A.; Varon, Y.; Parker, C.G.; Nelson, R.E.; B
Development 120, 1747-1758, 1994
A:Title: Tiggrin, a novel Drosophila extracellular matrix protein that functions as a
A:Reference number: Z17625; MUID:95009506; PMID:7924982
A:Accession: T13169
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2186 <FOG>
A:Cross-references: EMBL:U09506; NID:g493069; PID:g493070; PIDN:AAA56998.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0011722
C:Function:
A:Description: functions as a ligand for Drosophila alpha PS2 beta PS integrins

```

Query Match      8.7%; Score 9; DB 2; Length 2186;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 52 GSSASAS 60
Db 2069 GSSASAS 2077

RESULT 3
10.6K hypothetical protein F15D2.14 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: H86426
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86426
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-97 <STO>
A:CROSS-references: GB:AE005172; NID:g1120773; PIDN:AAG30954.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match      7.7%; Score 8; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 53 SSASASAS 60
Db 68 SSASASAS 75

RESULT 4
D96530
probable ribosomal protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96530
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-116 <STO>
A:CROSS-references: GB:AE005173; NID:g5430760; PIDN:AAD43160.1; GSPDB:GN00141
C:Genetics:
A:Gene: F13F21.17
A:Map position: 1

Query Match      7.7%; Score 8; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 54 SASASAS 61
Db 91 SASASAS 98

RESULT 5
T05954
transcription factor myb4 - barley (fragment)
C:Species: Hordeum vulgare (barley)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T05954
R:Gubler, F.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z15470
A:Accession: T05954
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-288 <GUB>
A:CROSS-references: EMBL:X99973; PIDN:CAA68235.1
A:Experimental source: cv. Himalaya, endosperm
C:Genetics:
A:Gene: myb4
C:Superfamily: Arabidopsis 28K leaf-specific myb-related protein; myb DNA-binding rep
C:Keywords: DNA binding; transcription regulation
F:60-110/Domain: myb DNA-binding repeat homology <MYB>

Query Match      7.7%; Score 8; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 53 SSASASAS 60
Db 197 SSASASAS 204

RESULT 6
T00500
probable elicitor response element-binding protein WRKY3 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 23-Mar-2001
C:Accession: T00500; C84623
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, November 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.
A:Reference number: Z14164
A:Accession: T00500
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-317 <ROU>
A:CROSS-references: EMBL:AC002391; NID:g2642427; PID:g2642432; GSPDB:GN00060; ATSP:T2
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84623
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-317 <STO>
A:CROSS-references: GB:AE002093; NID:g2642432; PIDN:AAB87100.1; GSPDB:GN00139
C:Genetics:
A:Gene: WRKY3; ATSP:T20D16.5; At2g23320
A:Map position: 2
A:Introns: 218/2; 260/2

Query Match      7.7%; Score 8; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 53 SSASASAS 60

```

Db 52 SSASASAS 59
|||||||

RESULT 7

T01597
Hypothetical protein At2g44700 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F16B22.19
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 16-Feb-2001
C:Accession: T01597; G84881
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F16B22 genomic sequence.
A:Reference number: Z14284
A:Accession: T01597
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-368 <ROU>
A:Cross-references: EMBL:AC003672; NID:g3341671; PID:g3341690
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84881
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <STO>
A:Cross-references: GB:AE002093; NID:g3341690; PIDN:AAC27472.1; GSPDB:GN00139
C:Genetics:
A:Gene: F16B22.19; At2g44700
A:Map position: 2
C:Superfamily: caffeoyl-CoA 3-O-methyltransferase

Query Match 7.7% Score 8; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
|||||||
Db 17 SSASASAS 24

RESULT 8

G70929
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022031; GB:AL123456; NID:g3250699; PIDN:CAA17711.1; PID:e125460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 7.7% Score 8; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GSSASASA 59

Db 156 GSSASASA 163
|||||||

RESULT 9

I38240
transcription factor SOX4 - human
N:Alternate names: sex-determining region Y box 4 protein
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 31-Jan-2000
C:Accession: I38240; S22938; S21479; S31724
R:Farr, C.J.; Easty, D.J.; Ragousis, J.; Collignon, J.; Lovell-Badge, R.; Goodfellow Mamm. Genome 4, 577-584, 1993
A:Title: Characterization and mapping of the human SOX4 gene.
A:Reference number: I38240; MUID:94093204; PMID:8268656
A:Accession: I38240
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-474 <RES>
A:Cross-references: EMBL:X70683; NID:g36552; PIDN:CAA50018.1; PID:g36553
R:Denny, P.; Swift, S.; Brand, N.; Dabhadre, N.; Barton, P.; Ashworth, A. Nucleic Acids Res. 20, 2887, 1992
A:Title: A conserved family of genes related to the testis determining gene, SRY.
A:Reference number: S22935; MUID:92310993; PMID:1614875
A:Accession: S22938
A:Molecule type: mRNA
A:Residues: 70,'P',72-123 <DEN>
A:Cross-references: EMBL:X65661; NID:g36551; PIDN:CAA46612.1; PID:g938230
C:Genetics:
A:Gene: GDB:SOX4
A:Cross-references: GDB:250365; OMIM:184430
A:Map position: 6p22.2-6p22.1p23-6p23
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: DNA binding; transcription regulation
F:56-131/Domain: HMG box homology <HMG1>

Query Match 7.7% Score 8; DB 2; Length 474;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
|||||||
Db 266 SASASASS 273

RESULT 10

D86280
protein T5E21.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D86280
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon ansen, N.F.; Hughes, B.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86280
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <STO>
A:Cross-references: GB:AE005172; NID:g7527718; PIDN:AAF63167.1; GSPDB:GN00141
C:Genetics:
A:Gene: T5E21.7
A:Map position: 1

Query Match 7.7% Score 8; DB 2; Length 514;
Best Local Similarity 100.0%; Pred. No. 3.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
 Db 201 SSASASAS 208
 |||||

RESULT 11
 B71807
 DNA repair protein(recombination protein n) - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
 C:Accession: B71807
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557; PMID:9923682
 A:Accession: B71807
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-522 <ARN>
 A:Cross-References: GB:AE001566; GB:AE001439; NID:g4156051; PIDN:AA007019.1; PID:g415608
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: recN
 C:Superfamily: recN protein

Query Match 7.7%; Score 8; DB 2; Length 522;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KDLERLN 74
 Db 185 KDLERLN 192
 |||||

RESULT 12
 A64694
 DNA repair protein - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 17-Mar-2000
 C:Accession: A64694
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
 Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467; PMID:9252185
 A:Accession: A64694
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-524 <TOM>
 A:Cross-References: GB:AE000639; GB:AE000511; NID:g2314560; PIDN:AA008433.1; PID:g231456
 C:Superfamily: recN protein

Query Match 7.7%; Score 8; DB 2; Length 524;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KDLERLN 74
 Db 185 KDLERLN 192
 |||||

RESULT 13
 E96492
 hypothetical protein F5A13.3 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C:Accession: E96492

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: E96492
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-538 <STO>
 A:Cross-References: GB:AE005173; NID:g9802762; PIDN:AAF99831.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F5A13.3
 A:Map position: 1
 C:Superfamily: Arabidopsis thaliana hypothetical protein T24H24.5

Query Match 7.7%; Score 8; DB 2; Length 538;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
 Db 476 SSASASAS 483
 |||||

RESULT 14
 AC3528
 extracellular serine proteinase (EC 3.4.21.-) [Imported] - Brucella melitensis (strai
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AC3528
 R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mufier, C.; Los, T.; Ivanov
 .; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AC3528
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-591 <KUR>
 A:Cross-References: GB:AE008918; PIDN:AA153390.1; PID:g17984284; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI10149
 A:Map position: II
 C:Keywords: hydrolase; serine proteinase

Query Match 7.7%; Score 8; DB 2; Length 591;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GSSASASA 59
 Db 261 GSSASASA 268
 |||||

RESULT 15
 S50591
 myb-related protein YER088c - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
 C:Accession: S50591
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmid 9747, 8198, 9781, and lambda clo
 A:Reference number: S50436
 A:Accession: S50591
 A:Molecule type: DNA

A:Residues: 1-670 <DIE>
A:Cross-references: EMBL:U18839; NID:g603313; PIDN:AAB64643.1; PID:g603326; GSPDB:GN0000
C:Genetics:
A:Gene: SGD:DOT6; MIPS:YER088c
A:Cross-references: SGD:S0000890
A:Map position: 5R
C:Superfamily: myb-related hypothetical protein YER088c; myb DNA-binding repeat homology
F:67-117/Domain: myb DNA-binding repeat homology <MYB>

Query Match 7.7% Score 8; DB 1; Length 670;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	54	SASASASS	61
Db	58	SASASASS	65

Search completed: July 16, 2003, 12:53:12
Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 18:40:05 ; Search time 171 Seconds
(without alignments)
1254.691 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 537

Sequence: 1 MQNSDYFYAQRCCQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09814661/runat_16072003_115311_14413/app_query.fasta_1.263
-DB=PublishedApplications_NA -OFT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdd -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09814661@cgn_1.1.80 @runat_16072003_115311_14413
-NCPU=6 -ICPU=3 -NO.MAP -LARGEOUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	537	100.0	1158	11	US-09-814-661A-1

c	2	78.5	14.6	2568	15	US-10-128-714-488	Sequence 488, App
c	3	77.5	14.6	2683	15 <td>US-10-128-714-5488</td> <td>Sequence 5488, Ap</td>	US-10-128-714-5488	Sequence 5488, Ap
c	4	74.5	14.4	2940917	15 <td>US-10-027-632-174763</td> <td>Sequence 174763,</td>	US-10-027-632-174763	Sequence 174763,
c	5	73.5	13.7	1278	11 <td>US-09-738-626-1124</td> <td>Sequence 1124, Ap</td>	US-09-738-626-1124	Sequence 1124, Ap
c	6	73.5	13.7	3309400	11 <td>US-09-738-626-1124</td> <td>Sequence 1124, Ap</td>	US-09-738-626-1124	Sequence 1124, Ap
c	7	71.5	13.3	1233	11 <td>US-09-738-626-1149</td> <td>Sequence 1149, Ap</td>	US-09-738-626-1149	Sequence 1149, Ap
c	8	71.5	13.3	3309400	11 <td>US-09-738-626-1149</td> <td>Sequence 1149, Ap</td>	US-09-738-626-1149	Sequence 1149, Ap
c	9	70.5	13.1	611	15 <td>US-10-124-557-27</td> <td>Sequence 27, Appl</td>	US-10-124-557-27	Sequence 27, Appl
c	10	70.5	13.1	1287	15 <td>US-10-156-761-3462</td> <td>Sequence 3462, Ap</td>	US-10-156-761-3462	Sequence 3462, Ap
c	11	70.5	13.1	5381	11 <td>US-09-751-877-4</td> <td>Sequence 4, Appli</td>	US-09-751-877-4	Sequence 4, Appli
c	12	70.5	13.1	5381	12 <td>US-09-842-364-4</td> <td>Sequence 4, Appli</td>	US-09-842-364-4	Sequence 4, Appli
c	13	70.5	13.1	81001	11 <td>US-09-751-877-1</td> <td>Sequence 1, Appli</td>	US-09-751-877-1	Sequence 1, Appli
c	14	70.5	13.1	81001	12 <td>US-09-842-364-1</td> <td>Sequence 1, Appli</td>	US-09-842-364-1	Sequence 1, Appli
c	15	70.5	13.1	9025608	15 <td>US-10-156-761-1</td> <td>Sequence 1, Appli</td>	US-10-156-761-1	Sequence 1, Appli
c	16	70	13.0	719	15 <td>US-10-027-632-144984</td> <td>Sequence 144984,</td>	US-10-027-632-144984	Sequence 144984,
c	17	70	13.0	111282	15 <td>US-10-094-989-3</td> <td>Sequence 3, Appli</td>	US-10-094-989-3	Sequence 3, Appli
c	18	69.5	12.9	5387	15 <td>US-10-001-873-22</td> <td>Sequence 22, Appl</td>	US-10-001-873-22	Sequence 22, Appl
c	19	69.5	12.9	536165	12 <td>US-09-939-964-1</td> <td>Sequence 1, Appli</td>	US-09-939-964-1	Sequence 1, Appli
c	20	69	12.8	978	10 <td>US-09-864-761-1564</td> <td>Sequence 1564, Ap</td>	US-09-864-761-1564	Sequence 1564, Ap
c	21	69	12.8	1425	11 <td>US-09-835-996A-3</td> <td>Sequence 3, Appli</td>	US-09-835-996A-3	Sequence 3, Appli
c	22	69	12.8	2128	11 <td>US-09-835-996A-42</td> <td>Sequence 42, Appl</td>	US-09-835-996A-42	Sequence 42, Appl
c	23	69	12.8	2128	15 <td>US-10-098-841-290</td> <td>Sequence 290, App</td>	US-10-098-841-290	Sequence 290, App
c	24	69	12.8	4826	15 <td>US-10-044-090-126</td> <td>Sequence 126, App</td>	US-10-044-090-126	Sequence 126, App
c	25	69	12.8	4826	15 <td>US-10-084-817-304</td> <td>Sequence 304, App</td>	US-10-084-817-304	Sequence 304, App
c	26	68.5	12.8	315	10 <td>US-09-923-876-3637</td> <td>Sequence 3637, Ap</td>	US-09-923-876-3637	Sequence 3637, Ap
c	27	68.5	12.8	2358	11 <td>US-09-801-368-347</td> <td>Sequence 347, App</td>	US-09-801-368-347	Sequence 347, App
c	28	68.5	12.8	3207	11 <td>US-09-801-368-241</td> <td>Sequence 241, App</td>	US-09-801-368-241	Sequence 241, App
c	29	68.5	12.8	3834	15 <td>US-10-128-714-7039</td> <td>Sequence 7039, Ap</td>	US-10-128-714-7039	Sequence 7039, Ap
c	30	68.5	12.8	7067	11 <td>US-09-764-877-3222</td> <td>Sequence 3222, Ap</td>	US-09-764-877-3222	Sequence 3222, Ap
c	31	68.5	12.8	17087	11 <td>US-09-070-927A-47</td> <td>Sequence 47, Appl</td>	US-09-070-927A-47	Sequence 47, Appl
c	32	68	12.7	1461	11 <td>US-09-801-368-397</td> <td>Sequence 397, App</td>	US-09-801-368-397	Sequence 397, App
c	33	68	12.7	3060	15 <td>US-10-270-333-92</td> <td>Sequence 92, Appl</td>	US-10-270-333-92	Sequence 92, Appl
c	34	68	12.7	5340	15 <td>US-10-270-333-91</td> <td>Sequence 91, Appl</td>	US-10-270-333-91	Sequence 91, Appl
c	35	68	12.7	7530	11 <td>US-09-969-708-211</td> <td>Sequence 211, App</td>	US-09-969-708-211	Sequence 211, App
c	36	68	12.7	7530	11 <td>US-09-880-107-3693</td> <td>Sequence 3693, Ap</td>	US-09-880-107-3693	Sequence 3693, Ap
c	37	68	12.7	10383	11 <td>US-09-960-253-181</td> <td>Sequence 181, App</td>	US-09-960-253-181	Sequence 181, App
c	38	68	12.7	32189	11 <td>US-09-764-877-3291</td> <td>Sequence 3291, Ap</td>	US-09-764-877-3291	Sequence 3291, Ap
c	39	67.5	12.6	420	15 <td>US-10-027-632-91876</td> <td>Sequence 91876, A</td>	US-10-027-632-91876	Sequence 91876, A
c	40	67.5	12.6	420	15 <td>US-10-027-632-91876</td> <td>Sequence 91876, A</td>	US-10-027-632-91876	Sequence 91876, A
c	41	67.5	12.6	627	15 <td>US-10-124-557-93</td> <td>Sequence 93, Appl</td>	US-10-124-557-93	Sequence 93, Appl
c	42	67.5	12.6	660	15 <td>US-10-124-557-95</td> <td>Sequence 95, Appl</td>	US-10-124-557-95	Sequence 95, Appl
c	43	67.5	12.6	780	15 <td>US-10-156-761-6305</td> <td>Sequence 6305, Ap</td>	US-10-156-761-6305	Sequence 6305, Ap
c	44	67.5	12.6	793	10 <td>US-09-770-445-839</td> <td>Sequence 839, App</td>	US-09-770-445-839	Sequence 839, App
c	45	67.5	12.6	1036	15 <td>US-10-180-375-111</td> <td>Sequence 111, App</td>	US-10-180-375-111	Sequence 111, App

ALIGNMENTS

RESULT 1
US-09-814-661A-1
; Sequence 1, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814.661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: S. Cerevisiae
US-09-814-661A-1

Alignment Scores:
Pred. No.: 2, 2e-63
Score: 537.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Length: 1158
Matches: 104
Conservative: 0
Mismatches: 0

```

Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-814-661A-1 (1-1158)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
DB 549 ATGCAAAATCCCAAGACTACTTTTACGCTCAAAATCGCTGCCAACAAACAGCCCT 608
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
DB 609 TCACATTCGTCACCGTACCATGCGGAATTTAGAAAGGGTCCCTTTGCCACCTATGGCT 668
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAla 60
DB 669 GAGGTTCCTATGTGCTACTCAAACTCCATGGGCGAGCTCGCTTCTGCCCTCCGCTTCT 728
QY 61 SerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAspHisAspMet 80
DB 729 TCATTAGAATGTGGAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 788
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlnLysValGlu 100
DB 789 ACAACAACAATAATTTGGTTCTGCGGAATAAAATCTATGTTCAACCAGGTAAGGTCGAG 848
QY 101 GluMetAspPhe 104
DB 849 GAATGGACTTC 860

```

```

RESULT 2
US-10-128-714-488/c
; Sequence 488, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 488
; LENGTH: 2568
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-488

Alignment Scores:
Pred. No.: 1.6 Length: 2568
Score: 78.50 Matches: 26
Percent Similarity: 46.25% Conservative: 11
Best Local Similarity: 32.50% Mismatches: 24
Query Match: 14.62% Indels: 19
DB: 15 Gaps: 4

US-09-814-661A-2 (1-104) x US-10-128-714-488 (1-2568)

```

```

QY 2 GlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaProSer 21
DB 245 CAGTATTCAACTGACTTTTCTTCACTCAATACGCTGTGAC----- 204
QY 22 ThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPro----- 38
DB 203 -----TTGTCAATGACCGAA-----CTCCCAACCCCTCCAGCATCCCGC 165
QY 39 ---MetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAla 57
DB 164 CAGAGCTCTGAATCACCAGCTTTCAGCTTAAGCAGTCT-----CACGAGCCGAGTCC 111
QY 58 SerAlaSerSerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAsp 77
DB 110 TCTGCCAAGCCCTTAGAATGCTGCACCATTCTCTGGAAGATATCTTTTACTCTCGAT 51

RESULT 3
US-10-128-714-5488/c
; Sequence 5488, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5488
; LENGTH: 2683
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5488

Alignment Scores:
Pred. No.: 1.7 Length: 2683
Score: 78.50 Matches: 26
Percent Similarity: 46.25% Conservative: 11
Best Local Similarity: 32.50% Mismatches: 24
Query Match: 14.62% Indels: 19
DB: 15 Gaps: 4

US-09-814-661A-2 (1-104) x US-10-128-714-5488 (1-2683)

QY 2 GlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaProSer 21
DB 360 CAGTATTCAACTGACTTTTCTTCACTCAATACGCTGTGAC----- 319
QY 22 ThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPro----- 38
DB 318 -----TTGTCAATGACCGAA-----CTCCCAACCCCTCCAGCATCCCGC 280
QY 39 ---MetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAla 57
DB 279 CAGAGCTCTGAATCACCAGCTTTCAGCTTAAGCAGTCT-----CACGAGCCGAGTCC 226
QY 58 SerAlaSerSerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAsp 77

```


Qy 69 ---LeuGluGluArgLeuAsnSerIleAspHisAspMetAsnAsnLysPheGlySer 87
Db 1083807 CCATTTAAACACGCTTTTCGACGATCTTGCCTCAGTAATCGAGGGGTCTGCTCGAGC 1083866
Qy 88 GlyCluLeuLysSerMetPheAsnGlnGly 97
Db 1083867 GGATAGTTGGAATCACTACTCAAGAGCGGG 1083896

RESULT 9
US-10-124-557-27
; Sequence 27, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124.557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 611 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1..333

SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-10-124-557-27

Alignment Scores:
Pred. No.: 2.83 Length: 611
Score: 70.50 Matches: 21
Percent Similarity: 45.16% Conservative: 21
Best Local Similarity: 22.58% Mismatches: 26
Query Match: 13.13% Indels: 25
DB: 15 Gaps: 3

US-09-814-661A-2 (1-104) x US-10-124-557-27 (1-611)
Qy 17 GlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPhe-----ArgArgVal 34
Db 291 AAGAAGACACCTCCACCTTCAGGAGCATCTCAAACCATCAATAACAACCAACGTTCA 350
Qy 35 ProLeuProPro-----MetAlaGlu 41
Db 351 CCCAAACCAACCAACAAAGAACACTAAGAAAGTTATAGAAATCAGAGGAATAACAGAA 410
Qy 42 ValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 61
Db 411 GAACATTCTGTCTTGAAATCAAGAGTCCTCCTCCTCCTCCTCCTCCTCCTCCTCTCT 470
Qy 62 LeuGluMetTrpGlu-----LysAspLeuGluGlu 71
Db 471 TCAACAATTTGGAAAAATCAAGTCTTCCTCAAAAATTCAGCTGCTAATAGAGAAATTACAGAAG 530
Qy 72 ArgLeuAsnSerIleAspHisAspMetAsnAsnLys 84
Db 531 AAACCTCAAGTAAAGATATAACAAGAACAGAACTTAA 569

RESULT 10
US-10-156-761-3462/c
; Sequence 3462, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE OF INVENTION: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3462
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1287)
US-10-156-761-3462

Alignment Scores:
Pred. No.: 7.73 Length: 1287
Score: 70.50 Matches: 22
Percent Similarity: 41.98% Conservative: 12
Best Local Similarity: 27.16% Mismatches: 30
Query Match: 13.13% Indels: 17
DB: 15 Gaps: 2

US-09-814-661A-2 (1-104) x US-10-156-761-3462 (1-1287)
Qy 12 AsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThr----- 27
Db 728 AGCCGCTGCCCCAGCACCGCCGCCACCACCGCCGCCAGCCAGCCGCGATCCGCAG 669
Qy 28 ---MetAlaGluPheArgValProLeuProProMetAlaGluValProMetLeuSer 46
Db 668 GCCACGGCGCACCGGTCCACACGCCCTCCGCGGGCGAGTTCGAGCCCGCGCATCAGC 609
Qy 47 ThrGlnAsnSerMetGlySerSerAlaSer----- 58

APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: APOLIPOPROTEIN A-IV-RELATED PROTEIN: POLYPEPTIDE, POLYNUCLEOTIDE
; FILE REFERENCE: SEQUENCES AND BIALLELIC MARKERS THEREOF.
; CURRENT APPLICATION NUMBER: US/09/842,364
; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 09/599,362
; PRIOR FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: PCT/IB99/02058
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 09/469/099
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: US 60/113,686
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 60/141,032
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 5381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..918
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 919..930
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 1442..1498
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 1613..1724
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 2243..3940
; OTHER INFORMATION: exon 4
; NAME/KEY: misc_feature
; LOCATION: 3941..5381
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 319
; OTHER INFORMATION: 17-42-319 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 3213
; OTHER INFORMATION: 17-41-250 : polymorphic base C or T
; NAME/KEY: conflict
; LOCATION: 1241
; OTHER INFORMATION: 17-39-343 : T in ref genbank AC007707
; NAME/KEY: conflict
; LOCATION: 1447
; OTHER INFORMATION: 17-40-202 : G in ref genbank AC007707
; NAME/KEY: primer_bind
; LOCATION: 1..11022
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 553..11575
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 899..11920
; OTHER INFORMATION: 17-39.pu
; NAME/KEY: primer_bind
; LOCATION: 1246..12267
; OTHER INFORMATION: 17-40.pu
; NAME/KEY: primer_bind
; LOCATION: 1441..12461
; OTHER INFORMATION: 17-39.rp complement
; NAME/KEY: primer_bind
; LOCATION: 1632..12651
; OTHER INFORMATION: 17-40.rp complement
; NAME/KEY: primer_bind
; LOCATION: 2964..13984

OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 3432..14454
OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 300..318
OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer_bind
; LOCATION: 320..338
OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer_bind
; LOCATION: 3194..3212
OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer_bind
; LOCATION: 3214..3232
OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: misc_binding
; LOCATION: 307..331
OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc_binding
; LOCATION: 3201..3225
OTHER INFORMATION: 17-41-250.probe
; US-09-842-364-4
Alignment Scores:
Pred. No.: 53.1 Length: 5381
Score: 70.50 Matches: 23
Percent Similarity: 45.95% Conservative: 11
Best Local Similarity: 31.08% Mismatches: 21
Query Match: 13.13% Indels: 19
DB: 12 Gaps: 4
US-09-814-661A-2 (1-104) x US-09-842-364-4 (1-5381)
QY 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro 37
Db 2109 CAAGAGCCCTCGACCTGGGG-----CCACGCC 2138
QY 38 ProMetAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAla 57
Db 2139 TCGGTCTTGGTTCCTCCACAGAGGATCAGTCCGCGATCTGTCCCTTCTCCAGGACCTGAAAGAC--- 2255
QY 58 SerAlaSerSerLeuGluMetTrp-----GluLysAspLeuGluArg 72
Db 2199 GGGGCTAGCAGTCTGACGCGCTGGATATCTGTCCCTTCTCCAGGACCTGAAAGAC--- 2255
QY 73 LeuAsnSerIleAspHisAspMetAsnAsn---AsnLysPhe 85
Db 2256 -----AGCCTTGAGCAAGACCTCAACAATATGAACAAGTTC 2291
RESULT 13
US-09-751-877-1
; Sequence 1, Application US/09751877
; Patent No. US20020142949A1
; GENERAL INFORMATION:
; APPLICANT: Yen, Frances
; APPLICANT: Denison, Blake
; APPLICANT: Bour, Barbara
; APPLICANT: Bihain, Bernard
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Ebbets-Reed, Dana
; APPLICANT: Salter-Cid, Luisa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
; FILE REFERENCE: 89.US3.REG
; CURRENT APPLICATION NUMBER: US/09/751,877
; CURRENT FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 81001
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10946..12946
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 12947..12958
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 13470..13526
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 13641..13752
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 14271..15968
; OTHER INFORMATION: exon 4
; NAME/KEY: misc_feature
; LOCATION: 15969..17969
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 1239
; OTHER INFORMATION: 20-828-311 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12347
; OTHER INFORMATION: 17-42-319 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1241
; OTHER INFORMATION: 17-41-250 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 42218
; OTHER INFORMATION: 20-841-149 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 43442
; OTHER INFORMATION: 20-842-115 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 77058
; OTHER INFORMATION: 20-853-415 : polymorphic base C or T
; NAME/KEY: primer_bind
; LOCATION: 929..949
; OTHER INFORMATION: 20-828.pu
; NAME/KEY: primer_bind
; LOCATION: 1357..1377
; OTHER INFORMATION: 20-828.rp complement
; NAME/KEY: primer_bind
; LOCATION: 12029..12050
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 12581..12603
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 14992..15012
; OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 15460..15482
; OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 42070..42090
; OTHER INFORMATION: 20-841.pu
; NAME/KEY: primer_bind
; LOCATION: 42572..42591
; OTHER INFORMATION: 20-841.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45328..45347
; OTHER INFORMATION: 20-842.pu
; NAME/KEY: primer_bind
; LOCATION: 45863..45883
; OTHER INFORMATION: 20-842.rp complement
; NAME/KEY: primer_bind
; LOCATION: 76544..76664
; OTHER INFORMATION: 20-853.pu
; NAME/KEY: primer_bind
; LOCATION: 77166..77185

```

```

; OTHER INFORMATION: 20-853.rp complement
; NAME/KEY: primer_bind
; LOCATION: 1220..1238
; OTHER INFORMATION: 20-828-311.mis
; NAME/KEY: primer_bind
; LOCATION: 1240..1258
; OTHER INFORMATION: 20-828-311.mis complement
; NAME/KEY: primer_bind
; LOCATION: 12328..12346
; OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer_bind
; LOCATION: 12348..12366
; OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15222..15240
; OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer_bind
; LOCATION: 15242..15260
; OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: primer_bind
; LOCATION: 42199..42217
; OTHER INFORMATION: 20-841-149.mis
; NAME/KEY: primer_bind
; LOCATION: 42219..42237
; OTHER INFORMATION: 20-841-149.mis complement
; NAME/KEY: primer_bind
; LOCATION: 45423..45441
; OTHER INFORMATION: 20-842-115.mis
; NAME/KEY: primer_bind
; LOCATION: 45443..45461
; OTHER INFORMATION: 20-842-115.mis complement
; NAME/KEY: primer_bind
; LOCATION: 77039..77057
; OTHER INFORMATION: 20-853-415.mis
; NAME/KEY: primer_bind
; LOCATION: 77059..77077
; OTHER INFORMATION: 20-853-415.mis complement
; NAME/KEY: misc_binding
; LOCATION: 1227..1251
; OTHER INFORMATION: 20-828-311.probe
; NAME/KEY: misc_binding
; LOCATION: 12335..12359
; OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc_binding
; LOCATION: 15229..15253
; OTHER INFORMATION: 17-41-250.probe
; NAME/KEY: misc_binding
; LOCATION: 42206..42230
; OTHER INFORMATION: 20-841-149.probe
; NAME/KEY: misc_binding
; LOCATION: 45430..45454
; OTHER INFORMATION: 20-842-115.probe
; NAME/KEY: misc_binding
; LOCATION: 77046..77070
; OTHER INFORMATION: 20-853-415.probe
; US-09-751-877-1

```

```

Alignment Scores:
Pred. No.: 2.05e+03 Length: 81001
Score: 70.50 Matches: 23
Percent Similarity: 45.95% Conservative: 11
Best Local Similarity: 31.08% Mismatches: 21
Query Match: 13.13% Indels: 19
DB: 11 Gaps: 4

```

US-09-814-661A-2 (1-104) x US-09-751-877-1 (1-81001)

```

Qy 18 GlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgValProLeuPro 37
Db 14137 CAAGAGCCCTCGACCCCTGGG-----CCACAGCCC 14166
Qy 38 PrometAlaGluValProMetLeuSerThrGlnAnSerMetGlySerAlaSerAla 57

```



```

; OTHER INFORMATION: 20-841-149 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 45442
; OTHER INFORMATION: 20-842-115 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 77058
; OTHER INFORMATION: 20-853-415 : polymorphic base C or T
; NAME/KEY: primer_bind
; LOCATION: 929..949
; OTHER INFORMATION: 20-828.pu
; NAME/KEY: primer_bind
; LOCATION: 1357..1377
; OTHER INFORMATION: 20-828.rp complement
; NAME/KEY: primer_bind
; LOCATION: 12029..12050
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 12581..12603
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 14992..15012
; OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 15460..15482
; OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 42070..42090
; OTHER INFORMATION: 20-841.pu
; NAME/KEY: primer_bind
; LOCATION: 42572..42591
; OTHER INFORMATION: 20-841.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45328..45347
; OTHER INFORMATION: 20-842.pu
; NAME/KEY: primer_bind
; LOCATION: 45863..45883
; OTHER INFORMATION: 20-842.rp complement
; NAME/KEY: primer_bind
; LOCATION: 76644..76664
; OTHER INFORMATION: 20-853.pu
; NAME/KEY: primer_bind
; LOCATION: 77166..77185
; OTHER INFORMATION: 20-853.rp complement
; NAME/KEY: primer_bind
; LOCATION: 1220..1238
; OTHER INFORMATION: 20-828-311.mis
; NAME/KEY: primer_bind
; LOCATION: 1240..1258
; OTHER INFORMATION: 20-828-311.mis complement
; NAME/KEY: primer_bind
; LOCATION: 12328..12346
; OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer_bind
; LOCATION: 12348..12366
; OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15222..15240
; OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer_bind
; LOCATION: 15242..15260
; OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: primer_bind
; LOCATION: 42199..42217
; OTHER INFORMATION: 20-841-149.mis
; NAME/KEY: primer_bind
; LOCATION: 42219..42237
; OTHER INFORMATION: 20-841-149.mis complement
; NAME/KEY: primer_bind
; LOCATION: 45423..45441
; OTHER INFORMATION: 20-842-115.mis
; NAME/KEY: primer_bind
; LOCATION: 45443..45461
; OTHER INFORMATION: 20-842-115.mis complement

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:36:11 ; Search time 78 Seconds
(without alignments)
274.729 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 537
Sequence: 1 MQNSQDYFYAQNRCQQQAP.....FGSGELKSMFNQGVEMDF 104

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	15.3	1665	5 Q9VQ19	Q9vq19 drosophila
2	82	15.3	1673	5 Q9BH66	Q9bh66 drosophila
3	81	15.1	494	5 Q9U2M8	Q9u2m8 caenorhabdi
4	80.5	15.0	931	5 Q16300	Q16300 caenorhabdi
5	80	14.9	1983	4 Q9Y4D6	Q9y4d6 homo sapien
6	78.5	14.6	451	16 Q9K741	Q9k741 bacillus ha
7	78	14.5	343	5 Q95Q03	Q95q03 caenorhabdi
8	77.5	14.4	418	16 Q55881	Q55881 synchocyst
9	73.5	13.7	798	2 Q9EV58	Q9ev58 spiroplasma
10	73.5	13.7	891	5 Q18430	Q18430 geodia cydo
11	73	13.6	347	10 Q9L049	Q9l049 oryza sativ
12	73	13.6	1028	5 Q9V3C4	Q9v3c4 drosophila
13	72.5	13.5	675	10 Q9S9V7	Q9s9v7 arabidopsis
14	72.5	13.5	814	13 Q8UDM2	Q8uum2 oryza lat
15	72	13.4	849	3 Q96W43	Q96wm3 cryptococcu
16	71.5	13.3	133	3 Q13577	Q13577 saccharomyc

17	71.5	13.3	328	5 Q9VE66	Q9ve66 drosophila
18	71.5	13.3	963	13 Q90624	Q90624 gallus gall
19	71.5	13.3	1769	5 Q16625	Q16625 caenorhabdi
20	71	13.2	372	5 Q9V4D0	Q9v4d0 drosophila
21	71	13.2	456	16 Q9Z9G3	Q9z9g3 chlamydia p
22	71	13.2	774	16 Q9K1Z4	Q9k1z4 chlamydia p
23	71	13.2	774	16 Q9JSK8	Q9jsk8 chlamydia p
24	71	13.2	811	16 Q9Z6K5	Q9z6k5 chlamydia p
25	71	13.2	1097	5 Q95U52	Q95u52 drosophila
26	71	13.2	1833	5 Q9VM67	Q9vm67 drosophila
27	70.5	13.1	541	16 Q8ZC85	Q8zcg5 yersinia pe
28	70.5	13.1	734	10 Q48620	Q48620 populus tri
29	70.5	13.1	1272	10 Q93VQ3	Q93vg3 oryza sativ
30	70	13.0	176	5 Q9V823	Q9v823 drosophila
31	70	13.0	208	10 Q94HL4	Q94hl4 oryza sativ
32	70	13.0	624	5 Q9W2T0	Q9w2t0 drosophila
33	69.5	12.9	756	13 Q9DGK0	Q9dgg0 xenopus lae
34	69.5	12.9	837	11 Q9QZ11	Q9qz11 mus musculu
35	69.5	12.9	1318	5 Q95PH4	Q95ph4 dictyosteli
36	69	12.8	159	5 Q9NLD3	Q9nld3 caenorhabdi
37	69	12.8	455	10 Q9FZE5	Q9fze5 arabidopsis
38	69	12.8	507	5 Q44885	Q44885 caenorhabdi
39	69	12.8	929	5 Q9BLX2	Q9blx2 dictyosteli
40	69	12.8	1270	4 Q96JH2	Q96jh2 homo sapien
41	68.5	12.8	182	13 Q9PWN5	Q9pwn5 gallus gall
42	68.5	12.8	501	5 Q93840	Q93840 caenorhabdi
43	68.5	12.8	503	2 Q54414	Q54414 salmonella
44	68.5	12.8	504	2 Q53991	Q53991 salmonella
45	68.5	12.8	504	2 Q54489	Q54489 salmonella

ALIGNMENTS

RESULT 1

Q9VQ19	PRELIMINARY;	PRT; 1665 AA.
ID	Q9VQ19	
AC	Q9VQ19;	
DT	01-MAY-2000 (TREMBlrel. 13, Created)	
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)	
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE	CG8817 protein.	
GN	LILLI OR CG8817.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
FC	STRAIN=BERKELEY;	
FX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Fosler K.C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodes A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	

```

Query Match      15.3% Score 82; DB 5; Length 1673;
Best Local Similarity 28.6%; Pred. No. 12;
.. Matches      28; Conservative 15; Mismatches 41; Indels 14; Gaps 3;

QY      10 AQRCCQQQAPSTLRTVTWAEFRVPLPPMAEVPMLSTQNSMGSSAS--ASASSLEWWEK 67
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      902 ARKREHSSOSSNGNTPT---KKVATPQLVAAPLAKPTNTAGSSDSDSSSAESSK 957
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      68 DLEERLNSIDHDHNN-----NKEFGSGELKSMENQG 97
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      958 SSSSSSSDDTETQNTNCRIVKLNKITGAVQKKALLGSG 995
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
Q9U2M8
ID Q9U2M8 PRELIMINARY; PRT; 494 AA.
AC Q9U2M8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE V37H9A.3 protein.
DE V37H9A.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

```

```

Query Match      15.1%; Score 81; DB 5; Length 494;
Best Local Similarity 25.0%; Pred. No. 3.5;
Matches 30; Conservative 24; Mismatches 40; Indels 26; Caps 5;

QY   8 FYAQNRCQQQAQPSTLRTVTMAEF-----RRVPLPMPAEVPMLSTQNSMG 52
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DB   182 FYAKSGDVEAAKKYL--VEMKGFDMIOAQAAGLPVSATKPDPQAQTASTTLPRIQ 239
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

QY   53 SSASASASSLE-----MWEKOLEERLASID-HDMNNKNFGS-GELKSFMNCGKV EMD 103
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DB   240 SAASSSTGVENRGEXLSLLEKTFLIEQVRS AETNQMRFTLCDGVDGKVRLFEGWGKVAKD 299
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

RESULT 4
O16300
ID OI6300 PRELIMINARY; PRT; 931 AA.
AC OI6300;
DT 01-JAN-1998 (TEMBLrel..05, Created)
DI 01-JAN-1998 (TEMBLrel..05, Last sequence update)
DT 01-DEC-2001 (TEMBLrel..19, Last annotation update)
DE Hypothetical 108.3 kDa protein.
GN t05C3_3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdiltida; Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2.
RX MEDLINE=99069613; PubMed=9851916;
```

```
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RL investigating biology. The C. elegans Sequencing Consortium.";
RN Science 282:2012-2018(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Blanchard M., Bradshaw H., Kramer J.;
RT "The sequence of C. elegans cosmid T05C3.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL EMBL; AF016428; AAB65360.1; -.
KW Hypothetical protein.
SQ SEQUENCE 931 AA; 108326 MW; 41FADA31F8703F85 CRC64;

Query Match 15.0%; Score 80.5; DB 5; Length 931;
Best Local Similarity 21.0%; Pred. No. 8.3;
Matches 22; Conservative 20; Mismatches 34; Indels 29; Gaps 2;

QY 27 TNAEFRRVPLPMA-----EVPMLSTONGSGSSASASLEWWEKDL 69
DB 119 SLIQIDTPMPPTARLNLQNSPEDKTENMDQETMLINTEPKTSDSTPTASYLRQWQDI 178
QY 70 EERLNSID-----HDMNNKFGSGELKSMFNOGKVEEM 102
DB 179 ERKVGRIEIVAEQYRMQKKDLGLEOKLDKNAEKSDVNQAKLES 223

RESULT 5
QY4D6 PRELIMINARY; PRT; 1983 AA.
AC QY4D6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE KIAA0624 protein (Fragment).
GN KIAA0624.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
FR EMBL; AB014524; BAA31599.1; -.
FT NON_TER 1
SQ SEQUENCE 1983 AA; 221956 MW; EE1676A3B8ED37F9 CRC64;

Query Match 14.9%; Score 80; DB 4; Length 1983;
Best Local Similarity 33.9%; Pred. No. 23;
Matches 41; Conservative 13; Mismatches 35; Indels 32; Gaps 9;

QY 2 QNSQDYFAQNRCQOQAFS--TLRTVTMAEF-----RRVPLPMAEVPMLSTONGSSA 55
DB 1708 ENSKDVTAQNLRVRESGASPTFTSLREAESDNORRLS-PPF----PLEPAQKSRVSP 1763
QY 56 SA-----SASSLEWEKD-----LEERLNSID-HDMNNKFGSGELKSMFNOGKVEEM 103
DB 1764 LASFLQQQRASASLE-WEPEPHLYRSKLSKLSNVH-----GDLRLKSHPPKVRERH 1813
QY 104 F 104
```

```
Db 1814 F 1814

RESULT 6
Q9K741 PRELIMINARY; PRT; 451 AA.
AC Q9K741;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3532.
GN BH3532.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001519; BAB07251.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 451 AA; 52337 MW; 8FC1DACF77DF8034 CRC64;

Query Match 14.6%; Score 78.5; DB 16; Length 451;
Best Local Similarity 35.9%; Pred. No. 5.7;
Matches 23; Conservative 12; Mismatches 24; Indels 5; Gaps 2;

QY 39 MAEVPMLSTONGSGSSASASASLEWWEKDL-ERLNSIDHDMNNKFGSGELKSMFNOG 97
DB 392 LAGITSROTQLSVLSYVDNVQNLDRIOKEGEINSFNRDMNSPFFGSGEL----DDG 447
QY 98 KVEE 101
DB 448 QEE 451

RESULT 7
Q95Q03 PRELIMINARY; PRT; 343 AA.
AC Q95Q03;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y66D12A.12 protein.
GN Y66D12A.12
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2].
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL161712; CAC70134.1; -.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_2.
```

```
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW DNA-binding; zinc-finger.
SQ SEQUENCE 343 AA; 37946 MW; E91F42FB5E72A210 CRC64;

Query Match 14.5%; Score 78; DB 5; Length 343;
Best Local Similarity 33.3%; Pred. No. 4.6;
Matches 32; Conservative 15; Mismatches 33; Indels 16; Gaps 5;

QY 15 QQQQAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSGSSASASSSLEMMWE----KDLE 70
DB QQQQQAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSGSSASASSSLEMMWE----KDLE 70
QY 167 QQQQAPVATPAPASWIDQ-----LAATPALLPFPMPSSSTFSASSSSSEMEOQTPESPSSLD 222
DB QQQQAPVATPAPASWIDQ-----LAATPALLPFPMPSSSTFSASSSSSEMEOQTPESPSSLD 222
QY 71 ERLNSIDHDMNNKFG-----SGELKSMFNGKQVEEMD 103
DB ERLNSIDHDMNNKFG-----SGELKSMFNGKQVEEMD 103
QY 223 TVLTSM---MNNNEEAASSTGEIKE-EEEEEEVD 253
DB TVLTSM---MNNNEEAASSTGEIKE-EEEEEEVD 253

RESULT 8
Q55881 PRELIMINARY; PRT; 418 AA.
AC Q55881;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein slr0108.
GN SLR0108.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64 to 92 of the genome.";
RL DNA Res. 2:153-166(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D64004; BAA10643.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 418 AA; 47079 MW; B62E91DFD65584A6 CRC64;

Query Match 14.4%; Score 77.5; DB 16; Length 418;
Best Local Similarity 23.8%; Pred. No. 6.6;
Matches 24; Conservative 14; Mismatches 26; Indels 37; Gaps 3;

QY 7 YFYAQRCCQQQAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSGSSA-----SASA 59
DB YFYAQRCCQQQAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSGSSA-----SASA 59
QY 16 FFWLHKSQESDHSPEL-----APLPDTLEAKLQSLNSLVTAYIKAIHSA 65
DB FFWLHKSQESDHSPEL-----APLPDTLEAKLQSLNSLVTAYIKAIHSA 65

QY 60 SLEMMWEKDE-----ERLNSIDHDM 80
DB SLEMMWEKDE-----ERLNSIDHDM 80
QY 66 DSLQAWQWNLAPNSLIVILANPVEPIAKLQDSLNSWDNDL 106
DB DSLQAWQWNLAPNSLIVILANPVEPIAKLQDSLNSWDNDL 106

RESULT 9
Q9EV58 PRELIMINARY; PRT; 798 AA.
AC Q9EV58;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW DNA-binding; zinc-finger.
SQ SEQUENCE 343 AA; 37946 MW; E91F42FB5E72A210 CRC64;

Query Match 13.7%; Score 73.5; DB 2; Length 798;
Best Local Similarity 28.2%; Pred. No. 37;
Matches 22; Conservative 15; Mismatches 32; Indels 9; Gaps 3;

QY 22 TLRVTMAEFRVRPLPPMAEVPMLSTQNSGSSASASSSLEMMWEKDLERLNSIDHDMN 81
DB TLRVTMAEFRVRPLPPMAEVPMLSTQNSGSSASASSSLEMMWEKDLERLNSIDHDMN 81
QY 9 TISTLTAS-----IPAPLLAAVPLTNTLTSSNNNDYLPVKQI-----NGVNNNINSITIDKN 60
DB TISTLTAS-----IPAPLLAAVPLTNTLTSSNNNDYLPVKQI-----NGVNNNINSITIDKN 60
QY 82 NN-KFGSGELKSMFNGKQ 98
DB NN-KFGSGELKSMFNGKQ 98
QY 61 NNLVFGTSEGAFVLKQGE 78
DB NNLVFGTSEGAFVLKQGE 78

RESULT 10
Q18430 PRELIMINARY; PRT; 891 AA.
AC Q18430;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Myosin II (Fragment).
RN MYO-2.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Astrophorida; Geodiidae; Geodia.
OX NCBI_TaxID=6047;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175357; PubMed=9023017;
RA Lorenz B., Bohnsack R., Gamulin V., Steffen R., Mueller W.B.G.;
RT "Regulation of motility of cells from marine sponges by calcium.";
RL Cell. Signal. 8:517-524(1996).
DR EMBL; X94984; CAA64440.1;
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
FT NON_TER 1
SQ SEQUENCE 891 AA; 102245 MW; E7F393568C3715D0 CRC64;

Query Match 13.7%; Score 73.5; DB 5; Length 891;
Best Local Similarity 29.3%; Pred. No. 42;
Matches 22; Conservative 16; Mismatches 32; Indels 5; Gaps 2;

QY 13 RCQQQQAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSGSSASASSSLEMMWEKDLER 72
DB RCQQQQAPSTLRTVTMAEFRVRPLPPMAEVPMLSTQNSGSSASASSSLEMMWEKDLER 72
QY 627 RCAHEQTSTRLREYEMAEAR-----RTAESERDELODEV-QSATSKANSLAEKRRVNR 681
DB RCAHEQTSTRLREYEMAEAR-----RTAESERDELODEV-QSATSKANSLAEKRRVNR 681
QY 73 LNSIDHDMNNKFGS 87
DB LNSIDHDMNNKFGS 87
QY 682 LSTLEEDLEEQMS 696
DB LSTLEEDLEEQMS 696

RESULT 11
Q9LD49
```



```
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Tln24.15 protein.
GN Tln24.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Washu;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Murray J., Langston Y., Clarke K., Drone K.;
RT "The sequence of A. thaliana Tln24.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF149413; AAD40140.1; -.
DR HSSP; P08631; IAD5.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 675 AA; 75367 MW; E4406763B268C712 CRC64;

Query Match 13.5%; Score 72.5; DB 10; Length 675;
Best Local Similarity 25.7%; Pred. No. 39;
Matches 18; Conservative 16; Mismatches 33; Indels 3; Gaps 2;

OY 2 QNSODYFAONRCQOQOAPSTLRVTMAEPRRVLPPMAEVPMLSTQNSMGSSASASASS 61
DB 239 QNSWDFYHGASSSEESIPQTKDIE-AEMRLKLELQTMIDMYSSACKKEALTAKRKANE 297

OY 62 LEMMEKDL EE 71
DB 298 LQWKK--IEE 305

RESULT 14
Q8UUM2
ID Q8UUM2 PRELIMINARY; PRT; 814 AA.
AC Q8UUM2;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE RING3 protein.
GN RING3.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD-RR;
RA Matsuo M.Y., Asakawa S., Shimizu N., Kimura H., Nonaka M.;
```

```
RT "Nucleotide Sequence of the MHC Class I Region of a Teleost, the
RL Medaka.";
RL Immunogenetics 0:0-0(2002);
DR EMBL; AB073376; BAB83842.1; -.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_2.
DR PROSITE; PS00014; BROMODOMAIN_2; 2.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
SQ SEQUENCE 814 AA; 90116 MW; C71293789354623D CRC64;

Query Match 13.5%; Score 72.5; DB 13; Length 814;
Best Local Similarity 33.8%; Pred. No. 48;
Matches 23; Conservative 9; Mismatches 25; Indels 11; Gaps 3;

OY 20 PSTLRTV---TMAEPRRVLPPMAEVPMLSTQNSMGSSASASASASASASASASAS 76
DB 707 PSTLRELERYVMTCLRKPP-----RKPVYSTKNSAGKSRELEALEKQM---ELER 758
OY 77 DHDMNNK 84
DB 759 SGQLNSGK 766

RESULT 15
Q96WM3
ID Q96WM3 PRELIMINARY; PRT; 849 AA.
AC Q96WM3;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Vacuolar (H+)-ATPase subunit.
GN VPH1.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=40410;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H99;
RA Erickson T., Liu L., Gueyikian A., Williamson P.R.;
RT "Multiple Virulence Factors of Cryptococcus neoformans are dependent
on CnvPH 1.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF254790; AAK81705.1; -.
DR InterPro; IPR002490; V_ATPase_sub116.
DR Pfam; PF01496; V_ATPase_sub_a; 1.
SQ SEQUENCE 849 AA; 95979 MW; F82DA118187A09BC CRC64;

Query Match 13.4%; Score 72; DB 3; Length 849;
Best Local Similarity 29.2%; Pred. No. 57;
Matches 28; Conservative 11; Mismatches 35; Indels 22; Gaps 3;

OY 18 QAPSTLRVTVMAEPRR-----VLPMPMAEVPMLSTQNSMGSSASASASAS 62
DB 52 QRPTPLRLRLAEWARLRFRSQTSLSPPLGVPLAAVPPFTT---VGPPRAQNYDEL 108
OY 63 EMWEKDLERLNSIDHDNNNNKFGSGELKSMFNOGK 98
DB 109 EEKLKEHERRLN---EMNKSWEELGRKRKSELENK 140

Search completed: July 16, 2003, 12:44:25
Job time : 81 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 17:48:36 ; Search time 1415 Seconds
(without alignments)
1190.340 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 537
Sequence: 1 MMSQDYFYAQRNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09814661/runat_16072003_115309_14280/app_query.fasta_1.263
-DB=EST_QPWT-fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn.1.1.2874.0runat_16072003_115309_14280 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	432	80.4	796	17	AQ875934
C 2	345	64.2	466	17	AQ501881
C 3	290	54.0	784	17	AQ875860
C 4	164.5	30.6	935	17	CNS06NOL
C 5	129.5	24.1	800	17	CNS06OBP
C 6	94	17.5	562	17	AQ500866
C 7	88	16.4	473	17	AQ891633
C 8	86	16.0	675	9	AA776943
C 9	84.5	15.7	822	17	BH591350
C 10	84	15.6	775	17	BH210250
C 11	81.5	15.2	685	13	BM650746
C 12	81	15.1	626	17	BH038249
C 13	80.5	15.0	468	13	BI748909
C 14	80.5	15.0	476	13	BI749020
C 15	80.5	15.0	536	13	BI396639
C 16	80	14.9	725	12	BG619299
C 17	79.5	14.8	493	10	AW186488
C 18	79.5	14.8	537	12	BE822121
C 19	79	14.7	729	9	AU119465
C 20	78.5	14.6	344	12	BG405357
C 21	78	14.5	424	9	AA396172
C 22	78	14.5	936	12	BF236780
C 23	77.5	14.4	609	17	AQ398448
C 24	77.5	14.4	820	13	BI107027
C 25	76.5	14.2	663	13	BI251950
C 26	76.5	14.2	879	12	BG650359
C 27	75.5	14.1	484	14	BM954389
C 28	75.5	14.1	837	13	BI144889
C 29	75.5	14.1	898	17	CNS022N3
C 30	75	14.0	253	10	AV337834
C 31	75	14.0	533	12	BG860390
C 32	75	14.0	597	14	BQ241166
C 33	75	14.0	711	13	BM588321
C 34	75	14.0	909	17	CNS070YL
C 35	74.5	13.9	705	13	BM638444
C 36	74.5	13.9	977	9	AL558422
C 37	74	13.8	735	14	BQ281954
C 38	74	13.8	1018	17	CNS077FB
C 39	73.5	13.7	425	9	AJ280350
C 40	73.5	13.7	522	12	BF478792
C 41	73.5	13.7	552	10	AW646666
C 42	73.5	13.7	611	10	BE494635
C 43	73.5	13.7	685	13	BI176842
C 44	73	13.6	295	17	BH204547
C 45	73	13.6	303	17	AZ967687

ALIGNMENTS

RESULT 1
AQ875934/c
LOCUS
DEFINITION
V131C10 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces cerevisiae genomic 5', DNA sequence.
ACCESSION
AQ875934
VERSION
AQ875934.1
KEYWORDS
GSS.
SOURCE
baker's yeast.
ORGANISM
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 796)
Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,

Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
Unpublished (1999)
Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mfn-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTCTTTGGAAGTAC
Class: transposon-tagged.
Location/Qualifiers

FEATURES

source
1..796
/organism="Saccharomyces cerevisiae"
/strain="Y2278 - S288C background, cir(0) rho(0)"
/db_xref="taxon:4932"
/clone_lib="mfn-3xHA/lacZ Insertion Library, strain Y2278"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mfn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 218 a 167 c 209 g 202 t
ORIGIN

Alignment Scores:
Pred. No.: 1,13e-40 Length: 796
Score: 432.00 Matches: 86
Percent Similarity: 96.70% Conservative: 2
Best Local Similarity: 94.51% Mismatches: 2
Query Match: 80.45% Indels: 1
DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AQ875934 (1-796)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 340 ATCCAAATTCACAGACTGCTTTACGCTCAAAATCGCTGCCAACACAAAGCCCT 281
QY 21 SerThrLeuArgThrValThrMetAlaGlu-PheArgArgValProLeuProMetal 40
Db 280 TCCACATTGCGTACCGTGACCATGTGGACATTTAGAGGGTGCCCTTGCCACCTATGCG 221
QY 40 aGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAla 60
Db 220 TGAGGCTCTATGTTGTCTACTCAGAACCTCCATGGGAGCTCGCTTCTGCGTCCGCTTC 161
QY 60 rSerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAspHisAspMe 80
Db 160 TTCATTAGAAATCGGAAAGGATTGGAGGAGACATCACTCTATCGATCATGACAT 101
QY 80 tAsnAsnAsnLysPheGlySerGlyGluLeu 90
Db 100 GAACAACAACAAATTTGGTTCTGGCGAACTA 70

RESULT 2
AQ501881/c 466 bp DNA linear GSS 29-APR-1999
LOCUS
DEFINITION
V13D1 mfn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', DNA sequence.
ACCESSION
AQ501881
VERSION
AQ501881.1 GI:4707531
KEYWORDS
GSS.
baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

REFERENCE

AUTHORS
Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., destages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
TITLE
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
Unpublished (1999)
Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mfn-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTCTTTGGAAGTAC
Class: transposon-tagged.
Location/Qualifiers

FEATURES

source
1..466
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mfn-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mfn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 131 a 92 c 120 g 119 t
ORIGIN

Alignment Scores:
Pred. No.: 9.22e-31 Length: 466
Score: 345.00 Matches: 68
Percent Similarity: 94.59% Conservative: 2
Best Local Similarity: 91.89% Mismatches: 4
Query Match: 64.25% Indels: 0
DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AQ501881 (1-466)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 267 ATGCAAAATTCACAGACTACTTTTACGCTCAAAATCGCTGCCAACACAAAGCCCT 208
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuProMetal 40
Db 207 TCCACATTGCGTACCGTGACCATGGCGAATTTAGAGGGTGCCCTTGCCACCTATGCT 148
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAla 60
Db 147 GAGGTTCTCTATGTTGTCTACTCAAAATCCATGGGAGCTCGCTTCTGCTCGCTTCT 88
QY 61 SerLeuGluMetTrpGluLysAspLeuGluArgLeuAsn 74
Db 87 TCATTAGAAATGTGGAAAGGGGTCTGACGCTCAGTGAAC 46

RESULT 3
AQ875860/c 784 bp DNA linear GSS 08-NOV-1999
LOCUS
DEFINITION
V130C9 mfn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces cerevisiae genomic 5', DNA sequence.
ACCESSION
AQ875860
VERSION
AQ875860.1 GI:6288104
KEYWORDS
GSS.
baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 784)

AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., Desanges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

JOURNAL Unpublished (1999)

COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumare@yale.edu
te of mtn-3xHA/lacZ insertion.
Seq primer: GGCCCTCTCTCTTGTGGAAGTAC
Class: transposon-tagged.

FEATURES Location/Qualifiers

1..784

/organism="Saccharomyces cerevisiae"

/strain="Y2278 - S288C background, cir(0) rho(0)"

/db_xref="taxon:4932"

/clone_lib="mtn-3xHA/lacZ Insertion Library, strain Y2278"

/lab_host="E. coli"

/note="Vector: pHS56-Sal; A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

source

1..784

/organism="Saccharomyces cerevisiae"

/strain="Y2278 - S288C background, cir(0) rho(0)"

/db_xref="taxon:4932"

/clone_lib="mtn-3xHA/lacZ Insertion Library, strain Y2278"

/lab_host="E. coli"

/note="Vector: pHS56-Sal; A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 236 a 147 c 202 g 199 t

ORIGIN

Alignment Scores:

Pred. No.: 6.24e-24 Length: 784

Score: 290.00 Matches: 60

Percent Similarity: 98.36% Conservative: 0

Best Local Similarity: 98.36% Mismatches: 0

Query Match: 54.00% Indels: 1

DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x A0875860 (1-784)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20

Db 247 ATGCAAAATTCACAGACTACTTTTACGCTCAAAATCGCTGCCAACACAAAGCCCT 188

QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProLeuPro-MetAla 40

Db 187 TCCACATTGCGPACCGTGACCATGCGGAATTTAGAAAGGTGCTTTGCTACCTATGCG 128

QY 40 aGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlase 60

Db 127 TGAGGTTCCTATGTGTCTACTCAAAACTCAATGCGACCTCCGCTTCTGCTCCGCTTC 68

QY 60 r 60

Db 67 T 67

RESULT 4

CNS06NOL 935 bp DNA linear GSS 17-JUN-2001

LOCUS T3 end of clone AU0AA008E08 of library AU0AA from strain CBS 3082

DEFINITION of Saccharomyces kluyveri, genomic survey sequence.

ACCESSION AL406171

VERSION AL406171.1

KEYWORDS GSS.

SOURCE Saccharomyces kluyveri.

ORGANISM Saccharomyces kluyveri

REFERENCE 1 (bases 1 to 935)

AUTHORS

Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Bottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durand, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekla, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Winkler, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 935)

AUTHORS Neuveglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F., Gaillardin, C. and Casaregola, S.

TITLE Genomic exploration of the hemiascomycetous yeasts: 9.

JOURNAL Saccharomyces kluyveri

MEDLINE FEBS Lett. 487 (1), 56-60 (2000)

PUBMED 20584719

REFERENCE 3 (bases 1 to 935)

AUTHORS Direct Submission

TITLE Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,

JOURNAL 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES Location/Qualifiers

1..935

/organism="Saccharomyces kluyveri"

/strain="CBS 3082"

/db_xref="taxon:4934"

/clone="AU0AA008E08"

/clone_lib="AU0AA"

/note="end : T3"

misc_feature

<6..>227

/note="similar to Saccharomyces cerevisiae ORF YLR437c [hypothetical protein]"

/evidence=not_experimental

BASE COUNT 270 a 186 c 214 g 261 t

ORIGIN 4 others

Alignment Scores:

Pred. No.: 5.2e-09 Length: 935

Score: 164.50 Matches: 40

Percent Similarity: 57.14% Conservative: 16

Best Local Similarity: 40.82% Mismatches: 25

Query Match: 30.63% Indels: 17

DB: 17 Gaps: 4

US-09-814-661A-2 (1-104) x CNS06NOL (1-935)

QY 12 AsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPhe 31

Db 93 AATGTCTGTGTGCAAGATAACTCT-----AACTTTACTATCCAGAAATAC 137

QY 32 ArgArgValProLeuProProMetAlaGluValProMetLeuSerThrGlnAsnSerMet 51

Db 138 AAACGGTTCACCTTCATCTGCAAGCAACCACTTGTGTAAACCAACGCACCTGTT 197

QY 52 GlySerSerAlaSerAlaSerAlaSerLeuGluMetTrpGluLysAspLeuGluGlu 71

Db 198 TCCAGCTCT-----AGCTCTCTGGAATGTGGGAAGACAAAGTTGGACGAG 242

QY 72 ArgLeuAsnSerIleAspHisAspMetAsnAsnLysPheGlySerGlyGluLeu --- 90

```

|||||
Db 243 AGTTGAACAATATGATAATGACATCATGAGAACAAAGCTCGGTGCTGGCGACTTCATG 302
Qy 91 -----LysSerMetPheAsnGlnGlyLysValGluGluMetAspPhe 104
Db 303 TTGGTGTGTACAAAGAGACATTT-----GGTGAATTGGAAGACTGGTGTATT 350

```

```

RESULT 5
CNS060BP 800 bp DNA linear GSS 04-JUL-2001
LOCUS T3 end of clone AV0AA005B06 of library AV0AA from strain CBS 379 of
Saccharomyces exiguus, genomic survey sequence.
AL407867
AL407867.1 GI:12173946
GSS.

```

```

Saccharomyces exiguus.
Saccharomyces exiguus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 800)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bollin,Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durand,P., Lepingle,A., Liorente,B.,
Maupertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies

```

```

FEMS Lett. 487 (1), 3-12 (2000)
20584711
MEDLINE 11152876
PUBMED
REFERENCE 2 (bases 1 to 800)
Gaillardin,C. and Casaregola,S.
Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,
Gaillardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 6.
Saccharomyces exiguus
FEMS Lett. 487 (1), 42-46 (2000)
20584716
MEDLINE 11152881
PUBMED
REFERENCE 3 (bases 1 to 800)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Crepeux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

```

```

Location/Qualifiers
1. .800
/organism="Saccharomyces exiguus"
/strain="CBS 379"
/db_xref="taxon:34358"
/clone="AV0AA005B06"
/clone_lib="AV0AA"
/note="end : T3"
<656..>784
misc_feature
[ note="similar to Saccharomyces cerevisiae ORF YML058w [
hypothetical protein ]"
/evidence=not_experimental
]

```

```

BASE COUNT 269 a 133 c 136 g 232 t 30 others
ORIGIN
Alignment Scores:
Pred. No.: 5.53e-05 Length: 800
Score: 129.50 Matches: 31

```

```

Percent Similarity: 54.02% Conservative: 16
Best Local Similarity: 35.63% Mismatches: 29
Query Match: 24.12% Indels: 11
Db: 17 Gaps: 3

```

```

US-09-814-661A-2 (1-104) x CNS060BP (1-800)

```

```

Qy 15 GlnGlnGlnGlnAlaProSerThrLeuArgThrValThrMetAlaGluPheArgVal 34
Db 554 AGACAAAGACAAAGCTTAAGCGGTCTCCAGAAAGTCGCTCAATCTTGAATTTTTRRAGAGTT 613
Qy 35 ProLeuProPromet---AlaGluValPrometLeuSerThrGlnAsnSerMetGlySer 53
Db 614 CCAATGCCAAAGATCAATGTTCCACCTCCAAATGTTAGAAAAAATGCTACTGCTGTTCT 673
Qy 54 SerAlaSerAlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluArgLeu 73
Db 674 TCA-----TCAACTTTAGAGATGTGGATGAAACGTTTAATCAAGATTA 718
Qy 74 AsnSerIleAspHisAspMetAsnAsnLysPheGlySerGlyCyluLeuLysSerMet 93
Db 719 ANTGANATTGAT-----AACAGATAAATAATCTGACTTATCTGTTTC 763
Qy 94 PheAsnGlnGlyLysValGlu 100
Db 764 TTCCTAAGGCTGAGATGGAA 784

```

RESULT 6

```

AQ500866/c 562 bp DNA linear GSS 29-APR-1999
LOCUS V31A2 mtn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', DNA sequence.
ACCESSION AQ500866
VERSION AQ500866.1 GI:4706516
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 562)
Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatidis,D., Jansen,R.,
Umansky,L., Heideman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)
Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mtn-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTCTTGTGGAAGTAC
Class: transposon-tagged.
Location/Qualifiers
1..562
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library"
/lab_host="E: coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS56-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mtn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

```

```

BASE COUNT 195 a 87 c 127 g 153 t
ORIGIN
Alignment Scores:

```

FEATURES

```

Source
1..562
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library"
/lab_host="E: coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS56-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mtn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

```

Alignment Scores:

Pred. No.: 0.5 Length: 562
 Score: 94.00 Matches: 20
 Percent Similarity: 85.19% Conservative: 3
 Best Local Similarity: 74.07% Mismatches: 3
 Query Match: 17.50% Indels: 1
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x A0500866 (1-562)

QY 79 AspMetAsnAsnAspGlySerGlyGluLeuLysSerMetPheAsn-GlnGlyLy 98
 DB 561 GACATGTACACACAAATATGTTGGCGCAACTCAATCTATGTTCAACCCGGGTAA 502
 QY 98 sValGluGluMetAspPhe 104
 DB 501 GGTCTGGACATGGACTTC 483

RESULT 7

AQ091633/c
 LOCUS HS_3142_A1_B08_MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3142 Col=15 Row=C, DNA sequence.
 ACCESSION AQ091633
 VERSION AQ091633.1 GI:6347823
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 473)
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3142 row: C column: 15
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 473.
 Location/Qualifiers

FEATURES

source
 1..473
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3142 Col=15 Row=C"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 99 a 134 c 116 g 123 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 1.97 Length: 473
 Score: 88.00 Matches: 16
 Percent Similarity: 69.44% Conservative: 9
 Best Local Similarity: 44.44% Mismatches: 11
 Query Match: 16.39% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x A0819133 (1-473)

QY 37 ProPrometAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerAlaSer 56

Db 108 CTCCAGTTCGAGGTACCATGTATGATACACACAGGGTCTCGGGGAGTTTCTAATAC 49
 QY 57 AlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArg 72
 Db 48 TCCTCTCTCTCTGAAGACACAGGCCTGGGAGATGAATTAGAGATAAGA 1

RESULT 8

AA776943
 LOCUS ac40h12.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
 DEFINITION IMAGE:858983 3' similar to 7:001120 GLUCOSE-6-PHOSPHATASE (HUMAN
); contains element MER22 repetitive element ;, mRNA sequence.
 ACCESSION AA776943
 VERSION AA776943.1 GI:2836274
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 675)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
 White, Y., Wylie, T., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
 WashU-NCI human EST project

TITLE

JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Possible reversed clone: polyt not found
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 310.
 Location/Qualifiers

FEATURES

source
 1..675
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:858983"
 /clone_lib="Stratagene hNT neuron (#937233)"
 /dev_stage="hNT neurons"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dt.
 Differentiated, post mitotic hNT neurons. Average insert
 size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
 GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTT 3'"
 BASE COUNT 122 a 196 c 190 g 167 t
 ORIGIN

Alignment Scores:

Pred. No.: 5.8 Length: 675
 Score: 86.00 Matches: 16
 Percent Similarity: 57.45% Conservative: 11
 Best Local Similarity: 34.04% Mismatches: 20
 Query Match: 16.01% Indels: 0
 DB: 9 Gaps: 0

US-09-814-661A-2 (1-104) x AA776943 (1-675)

QY 37 ProPrometAlaGluValProMetLeuSerThrGlnAsnSerMetGlySerAlaSer 56

Db 134 CTCCAGTTCGAGGTACCATGTATGATACACACAGGGTCTCGGGGAGTTTCTAATAC 193

QY 57 AlaSerAlaSerSerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIle 76

Db 194 TCCTCTCTCTCTGAAGACACAGGCCTGGGAGATGAATTCTCCAGACACCCCTCACTG 253

[illegible]

Query Match: 15.08% Indels: 8
DB: 17 Gaps: 4

US-09-814-661A-2 (1-104) x BH038249 (1-626)

QY 8 PheTyrAlaGlnAsnArgCysGlnGlnGlnAlaProSerThrLeuArgThrValThr 27
DB 195 TTCTGAATCTTAACATCTCTCAATGCTCAGGCGCTTCACACATCAGGACCTTTGGA 254

QY 28 MetAlaGluPheArgValProLeuProProMetAlaGluValProMetLeuSerThr 47
DB 255 ATGAACATCAT-----CCTCTCGCTCCAGCAAGTGCGGTGTTGAGTTAT 305

QY 48 GlnAsnSerMetGlySerAlaSerAlaSerAla-----SerSerLeuGluMetTrp 65
DB 306 GTGAGTCGCGTGCAATTCAGTACGACACACCTGCGCTTCTCCCTCTGTCATTATGG 365

QY 66 GluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMetAsnAsnLysPhe 85
DB 366 AAGAAGGCGTTCATCCCAAGG-----GCTGTTGATAGA---CTTACAAATGGAACACTG 416

QY 86 GlySerGlyGluLeuLysSerMetPheAsnGlnGly 97
DB 417 TCAAAATGCTCCCTCTTTGTCATATAGCAAGGG 452

RESULT 13
BI748909/c
LOCUS ro83f04.y1 Heterodera glycines J2 pAMPl v8 Chiapelli McCarter
DEFINITION Heterodera glycines cDNA 5', mRNA sequence.

ACCESSION BI748909
VERSION BI748909.1 GI:15770711
KEYWORDS EST.

SOURCE Heterodera glycines.
ORGANISM Heterodera glycines

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.

AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarelshvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)

COMMENT Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810

Email: est@wustl.edu

The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapelli@wustl.edu & jmccarte@wustl.edu) at

Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.

High quality sequence stop: 396.

Location/Qualifiers

1. .468

/organism="Heterodera glycines"

/db_xref="taxon:51029"

/clone_lib="Heterodera glycines J2 pAMPl v8 Chiapelli

McCarter"

/dev_stage="enriched for 2nd stage juveniles"

/lab_host="DHI08"

/note="Vector: pAMPl (Gibco); Site_1: NotI; Site_2: SalI;

The library was constructed by Brandi Chiapelli and Dr.

James McCarter at Washington University, St. Louis. The

cDNA was made by using Dynabead oligo-dT priming (Dynal).

PCR based library using a modified protocol from the

SMART PCR cDNA Synthesis Kit from Clontech. Directionally

cloned into the UDG sites of pAMPl. Nematodes are the OP50
strain developed by Dr. Charlie Opperman of North
Carolina State University (see Genetics, 146:1311-8, 1997
) . Frozen J2 nematodes were provided by Dr. Rick Davis
also of NCSU.

BASE COUNT 103 a 124 c 108 g 133 t
ORIGIN

Alignment Scores:
Pred. No.: 14.8 Length: 468
Score: 80.50 Matches: 33
Percent Similarity: 42.57% Conservative: 10
Best Local Similarity: 32.67% Mismatches: 22
Query Match: 14.99% Indels: 36
DB: 13 Gaps: 6

US-09-814-661A-2 (1-104) x BI748909 (1-468)

QY 10 AlaGlnAsnArg-----CysGlnGlnGlnGlnAla----- 19
DB 278 GCACACACGCTACTCATTCGCCAGCACAACAGCATTTCCATGCCAACATCGGCACACG 219

QY 20 -----ProSer-----ThrLeuArgThrValThrMetAlaGlu 30
DB 218 GTTTATAGTCAGTCCCGCTCCCATCGCACGCGTGGAAACGCGTGGCGCGCGAA 159

QY 31 PheArgArgValProLeuProProMet---AlaGluValProMetLeuSerThrGlnAsn 49
DB 158 TTTCCGGAATCCAGTTGCTCCCATTTGTTGTAAGTGCCTGCGCATGCCAACAGAAAC 99

QY 50 SerMetGlySerSerAlaSerAlaSerAlaSerSerLeuGluMetTrpGlu-LysAspLe 69
DB 98 AGT-----CAATGGCATCAGAAATTT 78

QY 69 uGluGluArgLeuAsnSerIleAspHisAspMetAsnAsnLysPheGlySerGlyG 89
DB 77 GGAGGAATTTGGCAATGGC-----AATGAAACAAAGGCATATAAAAGACGCGGAAA 24

QY 89 u 89
DB 23 G 23

RESULT 14
BI749020/c
LOCUS ro84h07.y1 Heterodera glycines J2 pAMPl v8 Chiapelli McCarter

DEFINITION Heterodera glycines cDNA 5', mRNA sequence.

ACCESSION BI749020
VERSION BI749020.1 GI:15770822
KEYWORDS EST.

SOURCE Heterodera glycines.
ORGANISM Heterodera glycines

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.

AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarelshvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)

COMMENT Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810

Email: est@wustl.edu

The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapelli@wustl.edu & jmccarte@wustl.edu) at

REFERENCE 1 (bases 1 to 536)
AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.

Year	Number of cases	Rate per 100,000
1990	1,000	1.0
1991	1,100	1.1
1992	1,200	1.2
1993	1,300	1.3
1994	1,400	1.4
1995	1,500	1.5
1996	1,600	1.6
1997	1,700	1.7
1998	1,800	1.8
1999	1,900	1.9
2000	2,000	2.0
2001	2,100	2.1
2002	2,200	2.2
2003	2,300	2.3
2004	2,400	2.4
2005	2,500	2.5
2006	2,600	2.6
2007	2,700	2.7
2008	2,800	2.8
2009	2,900	2.9
2010	3,000	3.0
2011	3,100	3.1
2012	3,200	3.2
2013	3,300	3.3
2014	3,400	3.4
2015	3,500	3.5
2016	3,600	3.6
2017	3,700	3.7
2018	3,800	3.8
2019	3,900	3.9
2020	4,000	4.0

Fri Jul 18 09:41:47 2003

us-09-814-661a-2_1.rst

Page 10

Db 13 G 13

Search completed: July 17, 2003, 19:00:10
Job time : 1421 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 19:54:00 ; Search time 1490 Seconds
(without alignments)
1130.423 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 104

Sequence: 1 MONSDYFYAQRCCQQQAP.....FCGSELKSMFNGKVEEMDF 104

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cpn2.1/USPTO.spool/US09814661/runat_16072003_115401_15636/app_query.fasta_1.263
-DB=EST-QFMT-fastap SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661.ecgn.1.1.2874@runat_16072003_115401_15636 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST : *
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	67	64.4	466	17	AQ501881	AQ501881 V13D1 mTn
C 2	48	46.2	796	17	AQ875934	AQ875934 V13C10 m
C 3	47	45.2	784	17	AQ875860	AQ875860 V13OC9 mT
C 4	10	9.6	356	10	AW934334	AW934334 EST360177
C 5	10	9.6	441	9	AU075401	AU075401 AU075401
C 6	10	9.6	456	9	AT003439	AT003439 AT003439
C 7	10	9.6	458	9	AI776067	AI776067 EST257263
C 8	10	9.6	465	14	C98970	C98970 C98970 Rice
C 9	10	9.6	465	14	C98971	C98971 C98971 Rice
C 10	10	9.6	474	14	BQ908903	BQ908903 T016H09 O
C 11	10	9.6	480	13	BI097287	BI097287 SCUMT93
C 12	10	9.6	494	13	BI813628	BI813628 L002E03 O
C 13	10	9.6	503	9	AI782032	AI782032 EST262911
C 14	10	9.6	511	9	AI776408	AI776408 EST257508
C 15	10	9.6	512	12	BG890753	BG890753 EST516604
C 16	10	9.6	516	9	AU056338	AU056338 AU056338
C 17	10	9.6	525	10	AW651362	AW651362 EST329816
C 18	10	9.6	526	13	BI812885	BI812885 G003E10 O
C 19	10	9.6	535	10	AW648997	AW648997 EST327451
C 20	10	9.6	535	13	BJ450473	BJ450473 BJ450473
C 21	10	9.6	539	10	BE459655	BE459655 EST414947
C 22	10	9.6	549	10	AW651197	AW651197 EST329651
C 23	10	9.6	555	13	BI179726	BI179726 EST520671
C 24	10	9.6	556	9	AI897354	AI897354 EST266797
C 25	10	9.6	562	17	AQ500866	AQ500866 V31A2 mTn
C 26	10	9.6	576	12	BG126079	BG126079 EST471725
C 27	10	9.6	593	9	AI482755	AI482755 EST242078
C 28	10	9.6	597	10	BE343914	BE343914 EST409076
C 29	10	9.6	600	10	AW650436	AW650436 EST328890
C 30	10	9.6	622	10	AW039032	AW039032 EST281005
C 31	10	9.6	637	10	AW219120	AW219120 EST301602
C 32	10	9.6	659	10	AW931489	AW931489 EST357332
C 33	10	9.6	673	13	BI928330	BI928330 EST548219
C 34	10	9.6	681	12	BI124243	BI124243 EST469877
C 35	10	9.6	692	13	BI928442	BI928442 EST548331
C 36	10	9.6	740	13	BI934910	BI934910 EST554799
C 37	10	9.6	743	13	BI924673	BI924673 EST544562
C 38	10	9.6	753	13	BI933801	BI933801 EST553690
C 39	10	9.6	756	12	BG590532	BG590532 EST498374
C 40	10	9.6	1046	17	CNS06Y4U	AL420580 T3 end of
C 41	9	8.7	145	12	BF895156	BF895156 RC3-MT013
C 42	9	8.7	167	14	BP005675	BP005675 BP005675
C 43	9	8.7	230	17	CNS00RGI	AL087072 Arabidops
C 44	9	8.7	236	17	AZ875012	AZ875012 ZMO189A22
C 45	9	8.7	250	14	D46245	D46245 RICS10787A

ALIGNMENTS

RESULT 1
AQ501881/c

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AQ501881 V13D1 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', DNA sequence.

AQ501881 genomic 5', DNA sequence.
AQ501881.1 GI:4707531
GSS.

baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 466)
Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,

Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
 Unpublished (1999)
 Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mTn-3xHA/lacZ insertion.
 Seq primer: GGCTTCTTCTTGTGAAGTAC
 Class: transposon-tagged.

JOURNAL
 COMMENT

FEATURES
 Source

1.466
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library"
 /lab_host="E. coli"
 /note="Vector: PHSS6-Sal: A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in PHSS6-Sal: genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."
 BASE COUNT 131 a 92 c 120 g 119 t 4 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1.4e-55 Length: 466
 Score: 67.00 Matches: 67
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 64.42% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661a-2 (1-104) x A0501881 (1-466)

OY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
 DB 267 ATGCAAAATTCACAGACTACTTTTACCTCAAAATCCCTGCCAACAACAGCCCT 208
 OY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
 DB 207 TCCACATTGCGTACCGTCACCATGCGGCAATTTAGAGGGTGCCCTTCCACCATATGCT 148
 OY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSer 60
 DB 147 GAGGTCTCTATGTTGTCTACTCAAACTCCATGGGACGCTCCGCTTCTCGCTTCT 88
 OY 61 SerLeuGluMetTrpGluLys 67
 DB 87 TCATTAGAAATGCGGAAG 67

RESULT 2
 A0875934/c
 LOCUS
 DEFINITION V131C10 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces cerevisiae genomic 5', DNA sequence.
 ACCESSION A0875934
 VERSION A0875934.1 GI:6288178
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 796)
 Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

TITLE
 JOURNAL
 COMMENT

Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
 Unpublished (1999)
 Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mTn-3xHA/lacZ insertion.
 Seq primer: GGCTTCTTCTTGTGAAGTAC
 Class: transposon-tagged.

FEATURES
 Source

1.796
 /organism="Saccharomyces cerevisiae"
 /strain="Y2278 - S288C background, cir(0) rho(0)"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"
 /lab_host="E. coli"
 /note="Vector: PHSS6-Sal: A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in PHSS6-Sal: genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."
 BASE COUNT 218 a 167 c 209 g 202 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9.55e-37 Length: 796
 Score: 48.00 Matches: 48
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 46.15% Indels: 0
 DB: 17 Gaps: 0

US-09-814-661a-2 (1-104) x A0875934 (1-796)

OY 43 ProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerSerLeu 62
 DB 213 CCTATGTTGTCTACTCAGAACTCCATGGCAGCTCCGCTTCTCGCTTCTCAATA 154
 OY 63 GluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAspHisAspMetAsn 82
 DB 153 GAATGTGGGAAGAGGATTGGAGAGAGACTCAACTCTATCGATCATGACATGACATGAC 94
 OY 83 AsnLysPheGlySerGlyGluLeu 90
 DB 93 AACAAATTTGTTCTTCTGCGAATA 70

RESULT 3
 A0875860/c

LOCUS
 DEFINITION V130C9 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces cerevisiae genomic 5', DNA sequence.
 ACCESSION A0875860
 VERSION A0875860.1 GI:6288104
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 784)
 Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mtN-3xHA/lacZ insertion.
Seq primer: GGCTTCCTTCTTTGGAAGTAC
Class: transposon-tagged.

FEATURES

source
1. .784
/organism="Saccharomyces cerevisiae"
/strain="Y2278 - S288C background, cir(0) rho(0)"
/db_xref="taxon:4932"
/clone_lib="mtN-3xHA/lacZ Insertion Library, strain Y2278"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtN-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 236 a 147 c 202 g 199 t

ORIGIN

Alignment Scores:
Pred. No.: 8,91e-36 Length: 784
Score: 47.00 Matches: 60
Percent Similarity: 98.36% Conservative: 0
Best Local Similarity: 98.36% Mismatches: 0
Query Match: 45.19% Indels: 1
DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x AQ875860 (1-784)

Qy 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 247 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGCTGCCAACCAACAGCCCT 188
Qy 21 SerThrLeuArgThrValThrMetAlaGluPheArgArgValProMetAl 40
Db 187 TCCACATTCGTACCGTGACCATGGCGGAATTTAGAGGGTGCTTTGCCCTATGTC 128
Qy 40 aGluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlase 60
Db 127 TGAGTTCCCTATGTTGTCTACTCAAACTCCATGGGCGAGCTCCGCTTCTCCCGCTTC 68
Qy 60 r 60
Db 67 T 67

RESULT 4
AW934334/c
LOCUS AW934334 356 bp mRNA linear EST 18-MAY-2001
DEFINITION EST360177 tomato fruit mature green, TAMU Lycopersicon esculentum
CDNA clone cLEF59J17 5', mRNA sequence.

ACCESSION AW934334
VERSION AW934334.1 GI:8109735
KEYWORDS EST.
SOURCE tomato.

ORGANISM

Lycopersicon esculentum
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE

AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

FEATURES

Location/Qualifiers
1. .356
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF59J17"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBlueScript SK(-); Site1: EcorI; Site2: XhoI; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 85 a 92 c 66 g 112 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 5.18 Length: 356
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 10 Gaps: 0

US-09-814-661A-2 (1-104) x AW934334 (1-356)

Qy 53 SerSerAlaSerAlaSerAlaSerSerLeu 62

Db 353 TCCTCGCGCTCGGCTCAGCTCATCACTC 324

RESULT 5

LOCUS AW975401/c
DEFINITION AW975401 Rice panicle at flowering stage Oryza sativa (japonica cultivar-group) cDNA clone E2964_42, mRNA sequence.
ACCESSION AW975401
VERSION AW975401.1 GI:5056022
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).

ORGANISM

Oryza sativa (japonica cultivar-group)
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle at flowering stage
JOURNAL Unpublished (1996)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@agr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'

FEATURES

Location/Qualifiers
1. .441
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E2964_42"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 100 a 135 c 110 g 94 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 6.62 Length: 441
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-2 (1-104) x AU075401 (1-441)

QY 52 GlySerSerAlaSerAlaSerAlaSer 61

Db 159 GGGTCGCGCGTCGGCGTCGGCGTCGTCG 130

RESULT 6

AT003439/c

LOCUS 456 bp mRNA linear EST 26-OCT-2000
DEFINITION Magnaporthe grisea infected rice cDNA library Oryza sativa
CDNA clone mgir3C18, mRNA sequence.

ACCESSION AT003439

VERSION AT003439.1 GI:11025057

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 456)

Kim.S., Park.J. and Lee.Y.

Magnaporthe grisea infected rice cDNA library, cDNA clones, 3' mRNA

sequence

Unpublished (2000)

Contact: Yong-Hwan Lee

Seoul National University

103, Seodun-dong, Suwon 441744, Korea

Tel: 82-31-290-2446

Fax: 82-31-294-5881

Email: yonglee@plaza.snu.ac.kr

Submitted through BRIC(Biological Research Information Center) of

Korea

URL: http://bric.postech.ac.kr/

GeneNuri No. KS103872.

FEATURES

Source

1..456
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="mgir3C18"
/clone_lib="Magnaporthe grisea infected rice cDNA library"
/tissue_type="leaves"
/dev_stage="80 h after inoculated with Magnaporthe grisea strain 70-15"

/note="Vector: lambda ZAP XAR; Site_1: EcoRI; Site_2: XhoI
; The cDNA library was constructed from rice leaves
heavily infected by M. grisea. About 10% of cDNA clones in
the library was fungal origin (M. grisea) and the 90% came
from rice; The library was constructed from leaves 80 h
after inoculated with Magnaporthe grisea strain 70-15"

BASE COUNT 110 a 143 c 99 g 103 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 6.88 Length: 456
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-2 (1-104) x AT003439 (1-456)

QY 52 GlySerSerAlaSerAlaSerAlaSer 61
Db 122 GGGTCGCGCGTCGGCGTCGGCGTCGTCG 93

RESULT 7

AT1776067/c

LOCUS 458 bp mRNA linear EST 18-MAY-2001

DEFINITION EST257263 tomato resistant, Cornell Lycopersicon esculentum CDNA

clone cLER17L5, mRNA sequence.

ACCESSION AT1776067

VERSION AT1776067.1 GI:5274204

KEYWORDS EST.

SOURCE tomato.

ORGANISM

Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 458)

D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,

Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman

,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,

Giovannoni,J.J. and Martin,G.B.

Generation of ESTs from Pseudomonas resistant tomato

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA.

Email: http://www.genome.clemson.edu/orders/index.html

5 prime sequence.

FEATURES

Source

1..458
/organism="Lycopersicon esculentum"
/cultivar="Ril-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="cLER17L5"
/clone_lib="tomato resistant, Cornell"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"

/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
Directionally cloned cDNAs inserted into pBlueScript SK(-
) at 5' end with EcoRI and 3' end with XhoI site."

BASE COUNT 123 a 99 c 100 g 134 t 2 others

ORIGIN

Alignment Scores:
Pred. No.: 6.91 Length: 458
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-2 (1-104) x AT1776067 (1-458)

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62

Db 311 TCCTCTCGCTCAGCCTCAGCCTCATCATC 282

RESULT 8

C98970/c

LOCUS 465 bp mRNA linear EST 04-APR-2002

DEFINITION C98970 Rice panicle at flowering stage Oryza sativa (Japonica

cultivar-group) cDNA clone E3320_6A, mRNA sequence.

ACCESSION C98970

VERSION C98970.1 GI:3761722

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group).

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
AUTHORS 1 (bases 1 to 465)
TITLE /clone_lib="Rice panicle at flowering stage"
JOURNAL /dev_stage="flowering stage"
COMMENT /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

POLYA=NO. Location/Qualifiers
source
1. .465
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E3320_6A"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 100 a 148 c 113 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 7.04 Length: 465
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x C98970 (1-465)

QY 52 GlySerSerAlaSerAlaSer 61
|||||
Db 154 GGTCGTCGGCGTCGGCGTCGTCG 125

RESULT 9
C98971/c
LOCUS
DEFINITION C98971 Rice panicle at flowering stage Oryza sativa (japonica cultivar-group) cDNA clone E3320_82, mRNA sequence.
ACCESSION C98971.1 GI:3761723
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group).
Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 465)
Dong, H.T., Li, D.B., Zhuang, X.F., Dai, C.G., Sun, L.X., Pei, Y.X., Wu, H.F., Jiang, Y.X., Yu, F.C., Gao, Q.K. and Lou, Y.C.
A Gene Expression Screen in Oryza sativa
Unpublished (2001)
Contact: Dong HT
Laboratory of Functional Genetics
Bio-technology Institute of Zhejiang University
Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
Tel: 0086-571-86892051
Fax: 0086-571-86961525
Email: htdong@zjuem.zju.edu.cn
Seq primer: M13 forward primer.
Location/Qualifiers
source
1. .474
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="T016H09"
/clone_lib="Oryza sativa mature leaf library induced by M. grisea"
/tissue_type="leaf"
/dev_stage="Mature stage"
/note="Vector: pSport2"

BASE COUNT 116 a 146 c 108 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 7.19 Length: 474
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x BQ908903 (1-474)

QY 52 GlySerSerAlaSerAlaSer 61
|||||
Db 154 GGTCGTCGGCGTCGGCGTCGTCG 125

REFERENCE
AUTHORS 1 (bases 1 to 465)
TITLE /clone_lib="Rice panicle at flowering stage"
JOURNAL /dev_stage="flowering stage"
COMMENT /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

POLYA=NO. Location/Qualifiers
source
1. .465
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E3320_6A"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 100 a 148 c 113 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 7.04 Length: 465
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x C98970 (1-465)

QY 52 GlySerSerAlaSerAlaSer 61
|||||
Db 154 GGTCGTCGGCGTCGGCGTCGTCG 125

RESULT 9
C98971/c
LOCUS
DEFINITION C98971 Rice panicle at flowering stage Oryza sativa (japonica cultivar-group) cDNA clone E3320_82, mRNA sequence.
ACCESSION C98971.1 GI:3761723
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group).
Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 465)
Dong, H.T., Li, D.B., Zhuang, X.F., Dai, C.G., Sun, L.X., Pei, Y.X., Wu, H.F., Jiang, Y.X., Yu, F.C., Gao, Q.K. and Lou, Y.C.
A Gene Expression Screen in Oryza sativa
Unpublished (2001)
Contact: Dong HT
Laboratory of Functional Genetics
Bio-technology Institute of Zhejiang University
Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
Tel: 0086-571-86892051
Fax: 0086-571-86961525
Email: htdong@zjuem.zju.edu.cn
Seq primer: M13 forward primer.
Location/Qualifiers
source
1. .474
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="T016H09"
/clone_lib="Oryza sativa mature leaf library induced by M. grisea"
/tissue_type="leaf"
/dev_stage="Mature stage"
/note="Vector: pSport2"

BASE COUNT 116 a 146 c 108 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 7.19 Length: 474
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x BQ908903 (1-474)

QY 52 GlySerSerAlaSerAlaSer 61
|||||
Db 154 GGTCGTCGGCGTCGGCGTCGTCG 125

```

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
AUTHORS 1 (bases 1 to 465)
TITLE /clone_lib="Rice panicle at flowering stage"
JOURNAL /dev_stage="flowering stage"
COMMENT /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

POLYA=NO. Location/Qualifiers
source
1. .465
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E3320_6A"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 110 a 146 c 104 g 105 t
ORIGIN

Alignment Scores:
Pred. No.: 7.04 Length: 465
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x C98971 (1-465)

QY 52 GlySerSerAlaSerAlaSer 61
|||||
Db 136 GGTCGTCGGCGTCGGCGTCGTCG 107

RESULT 10
BQ908903/c
LOCUS
DEFINITION BQ908903 Oryza sativa mature leaf library induced by M. grisea Oryza sativa cDNA clone T016H09, mRNA sequence.
ACCESSION BQ908903
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 474)
Dong, H.T., Li, D.B., Zhuang, X.F., Dai, C.G., Sun, L.X., Pei, Y.X., Wu, H.F., Jiang, Y.X., Yu, F.C., Gao, Q.K. and Lou, Y.C.
A Gene Expression Screen in Oryza sativa
Unpublished (2001)
Contact: Dong HT
Laboratory of Functional Genetics
Bio-technology Institute of Zhejiang University
Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
Tel: 0086-571-86892051
Fax: 0086-571-86961525
Email: htdong@zjuem.zju.edu.cn
Seq primer: M13 forward primer.
Location/Qualifiers
source
1. .474
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="T016H09"
/clone_lib="Oryza sativa mature leaf library induced by M. grisea"
/tissue_type="leaf"
/dev_stage="Mature stage"
/note="Vector: pSport2"

BASE COUNT 116 a 146 c 108 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 7.19 Length: 474
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x BQ908903 (1-474)

QY 52 GlySerSerAlaSerAlaSer 61
|||||
Db 136 GGTCGTCGGCGTCGGCGTCGTCG 107

```

```

Db      146 GGGTCGTGGCGTCGGCGTCGGCGTCGTCG 117
RESULT 11
BI097287
LOCUS   BI097287                480 bp    mRNA    linear    EST 25-JUN-2001
DEFINITION SCUMTT93 Melaleuca alternifolia (Cheel) mRNA Melaleuca alternifolia
cDNA, mRNA sequence.
ACCESSION BI097287
VERSION   BI097287.1 GI:14548944
KEYWORDS EST.
SOURCE    tea tree.
ORGANISM  Melaleuca alternifolia
REFERENCE
AUTHORS   Shelton,D., Leach,D., Baverstock,P and Henry,R.
TITLE     Isolation of genes involved in secondary metabolism from Melaleuca
JOURNAL   alternifolia (Cheel) using expressed sequence tags (ESTs)
COMMENT   Plant Sci. 162 (1), 9-15 (2002)
          Contact: Shelton D
          Centre for Plant Conservation Genetics
          Southern Cross University
          P.O. Box 157, Lismore, NSW 2480, Australia
          Tel: 61 2 6620 3173
          Fax: 61 2 6622 2080
          Email: dshe1t10@scu.edu.au.
FEATURES             Location/Qualifiers
     source          1..480
                     /organism="Melaleuca alternifolia"
                     /db_xref="taxon:164405"
                     /clone_lib="Melaleuca alternifolia (Cheel) mRNA"
BASE COUNT  92 a  201 c  89 g  95 t  3 others
ORIGIN
Alignment Scores:
Pred. No.:      7.3      Length:      480
Score:          10.00    Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:     9.62%      Indels:      0
DB:              13        Gaps:         0

US-09-814-661A-2 (1-104) x BI097287 (1-480)

QY      52 GlySerSerAlaSerAlaSerAlaSer 61
       |||||||
Db      95 GGCCTCAAGCGCTCTCTCATCGGCGCTCTCA 124

RESULT 12
BI813628/c
LOCUS   BI813628                494 bp    mRNA    linear    EST 01-NOV-2001
DEFINITION L002E03 Oryza sativa mature leaf library induced by M.grisea Oryza
sativa cDNA clone L002E03, mRNA sequence.
ACCESSION BI813628
VERSION   BI813628.1 GI:16579533
KEYWORDS EST.
SOURCE    Oryza sativa.
ORGANISM  Oryza sativa
REFERENCE
AUTHORS   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X., Wu
A Gene Expression Screen in Oryza sativa
Unpublished (2001)
Contact: Dong HT
Laboratory of Functional Genetics
Bio-Technology Institute of Zhejiang University
Kaiyuan Road 268# Hangzhou, Zhejiang, P.R.China
Tel: 0086-571-86882051
Fax: 0086-571-86961525

Email: htdong@zjuem.zju.edu.cn
Seq primer: M13 forward primer.
FEATURES             Location/Qualifiers
     source          1..494
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /clone_lib="L002E03"
                     /clone_lib="Oryza sativa mature leaf library induced by
                     M.grisea"
                     /tissue_type="leaf"
                     /dev_stage="Mature stage"
                     /note="Vector: pSport2"
BASE COUNT  112 a  147 c  136 g  99 t
ORIGIN
Alignment Scores:
Pred. No.:      7.54      Length:      494
Score:          10.00    Matches:      10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:     9.62%      Indels:      0
DB:              13        Gaps:         0

US-09-814-661A-2 (1-104) x BI813628 (1-494)

QY      52 GlySerSerAlaSerAlaSerAlaSer 61
       |||||||
Db      218 GGGTCGTGGCGTCGGCGTCGGCGTCGTCG 189

RESULT 13
AI782032/c
LOCUS   AI782032                503 bp    mRNA    linear    EST 18-MAY-2001
DEFINITION EST262911 tomato susceptible, Cornell Lycopersicon esculentum cDNA
clone cLES17L20, mRNA sequence.
ACCESSION AI782032
VERSION   AI782032.1 GI:52800073
KEYWORDS EST.
SOURCE    tomato.
ORGANISM  Lycopersicon esculentum
REFERENCE 1 (bases 1 to 503)
AUTHORS   D' Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
Giovannoni,J.J. and Martin,G.B.
Generation of ESTs from Pseudomonas susceptible tomato
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES             Location/Qualifiers
     source          1..503
                     /organism="Lycopersicon esculentum"
                     /cultivar="Ril-13 (Rio Grande x Money Maker)"
                     /db_xref="taxon:4081"
                     /clone="cLES17L20"
                     /clone_lib="tomato susceptible, Cornell"
                     /tissue_type="leaf"
                     /dev_stage="4-week old"
                     /lab_host="SOLR"
XhoI: cLES - Tomato Pseudomonas Susceptible EST Library.
Directionally cloned cDNAs inserted into pBlueScript SK(-
) at 5' end with EcoRI and 3' end with XhoI site"
BASE COUNT  121 a  121 c  101 g  160 t
ORIGIN

```


Alignment Scores:
 Pred. No.: 7.7 Length: 503
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.62% Indels: 0
 DB: 9 Gaps: 0

US-09-814-661A-2 (1-104) x AI782032 (1-503)

Qy 53 SerSerAlaSerAlaSerAlaSerLeu 62
 Db 391 TCCTCTGCCTCAGCCTCAGCCTCATCACTC 362

RESULT 14

AI776408/c 511 bp mRNA linear EST 18-MAY-2001
 LOCUS
 DEFINITION EST257508 tomato resistant, Cornell Lycopersicon esculentum cDNA

ACCESSION AI776408
 VERSION AI776408.1 GI:5274449
 KEYWORDS EST.
 SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatoxophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

AUTHORS D'Ascenzo,M., Her,X., Lyman,J., Matern,A.L., Viston,T., Holt,I.E.,
 Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
 C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
 Giovannoni,J.J. and Martin,G.B.

TITLE Generation of ESTs from Pseudomonas resistant tomato
 JOURNAL Unpublished (1999)
 COMMENT Contact: CUGI

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source Location/Qualifiers

1..511
 /organism="Lycopersicon esculentum"
 /cultivar="Ril-12 (35S::Pto in Rio Grande x Money Maker)"
 /db_xref="taxon:4081"
 /clone="cLER18K12"
 /clone_lib="tomato resistant, Cornell"
 /tissue_type="leaf"
 /dev_stage="4-week old"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; cLER - Tomato Pseudomonas Resistant EST Library. SK(-)
 Directionally cloned cDNAs inserted into pBluescript SK(-)
) at 5' end with EcoRI and 3' end with XhoI site."
 BASE COUNT 143 a 88 c 153 g 127 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7.84 Length: 511
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.62% Indels: 0
 DB: 9 Gaps: 0

US-09-814-661A-2 (1-104) x AI776408 (1-511)

Qy 53 SerSerAlaSerAlaSerAlaSerLeu 62
 Db 113 TCCTCTGCCTCAGCCTCAGCCTCATCACTC 84

RESULT 15

BG890753/c

LOCUS
 DEFINITION EST516604 cSTD Solanum tuberosum cDNA clone cSTD19J9 5' sequence,
 mRNA sequence.

ACCESSION BG890753
 VERSION BG890753.1 GI:14267886
 KEYWORDS EST.
 SOURCE potato.

ORGANISM Solanum tuberosum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatoxophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 512)

AUTHORS van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemingo,A.,
 Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
 TITLE Generations of ESTs from dormant potato tubers
 JOURNAL Unpublished (2001)
 COMMENT Contact: Cathy Ronning

The Institute for Genomic Research
 For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com
 Seq primer: M13F-R.

FEATURES
 source Location/Qualifiers

1..512
 /organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="cSTD19J9"
 /clone_lib="cSTD"
 /tissue_type="dormant tuber"
 /dev_stage="one month post-harvest"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; This library targets genes expressed in dormant
 tubers. This library was made from sections of dormant
 tuber, avoiding the buds and epidermis. Tubers were stored
 for one month post-harvest at 40C. The tuber was peeled,
 well away from the surface. Then it was chopped into 1-2
 mm cubes and immediately frozen in liquid nitrogen. This
 library is noted as P4 in Tanksley lab notebooks."
 BASE COUNT 130 a 115 c 109 g 158 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7.85 Length: 512
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.62% Indels: 0
 DB: 12 Gaps: 0

US-09-814-661A-2 (1-104) x BG890753 (1-512)

Qy 53 SerSerAlaSerAlaSerAlaSerLeu 62
 Db 362 TCCTCCGCCTCGGCCTCAGCCTCATCACTC 333

Search completed: July 17, 2003, 20:49:00
 Job time : 1517 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: July 16, 2003, 12:48:22 ; Search time 29 seconds
(without alignments)
738.928 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 104
Sequence: 1 MNSODYFAQNRCCQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rviris.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	8.7	131	10 Q9LSX0	Q9LSX0 arabidopsis
2	9	8.7	406	5 Q9W404	Q9W404 drosophila
3	9	8.7	554	10 Q9LRF24	Q9LRF24 arabidopsis
4	9	8.7	588	11 Q9DC19	Q9DC19 mus musculus
5	9	8.7	877	5 Q24191	Q24191 drosophila
6	9	8.7	877	5 Q9VHA0	Q9VHA0 drosophila
7	9	8.7	2186	5 Q23984	Q23984 drosophila
8	9	8.7	2188	5 Q9VMD9	Q9VMD9 drosophila
9	8	7.7	97	10 Q9C754	Q9C754 arabidopsis
10	8	7.7	116	10 Q9XIA6	Q9XIA6 arabidopsis
11	8	7.7	121	5 Q9V5U8	Q9V5U8 drosophila
12	8	7.7	156	5 Q9VJV7	Q9VJV7 drosophila
13	8	7.7	208	5 Q9W102	Q9W102 drosophila
14	8	7.7	256	16 Q9KYH7	Q9KYH7 streptomyces
15	8	7.7	279	5 Q9N7U4	Q9N7U4 leishmania
16	8	7.7	288	10 Q96463	Q96463 hordeum vul

17	7.7	368	10	O80509	O80509 arabidopsis
18	7.7	399	10	O9LWJ8	O9LWJ8 oryza sativ
19	7.7	411	16	O53939	O53939 mycobacteri
20	7.7	468	10	O94JZ8	O94JZ8 arabidopsis
21	7.7	514	10	O9MA26	O9MA26 arabidopsis
22	7.7	521	5	O960Q1	O960Q1 drosophila
23	7.7	531	5	O9N8Y4	O9N8Y4 trypanosoma
24	7.7	538	10	O9FZH1	O9FZH1 arabidopsis
25	7.7	591	16	O8YDM6	O8YDM6 bruceella me
26	7.7	620	5	O9VKL0	O9VKL0 drosophila
27	7.7	636	11	O9Z2E2	O9Z2E2 mus musculu
28	7.7	664	5	O95ST4	O95ST4 drosophila
29	7.7	738	5	O02402	O02402 pinctada fu
30	7.7	755	5	O9N4K2	O9N4K2 caenorhabdi
31	7.7	763	10	O941Z6	O941Z6 oryza sativ
32	7.7	844	5	O9VL72	O9VL72 drosophila
33	7.7	931	10	O8S4P6	O8S4P6 zea mays (m
34	7.7	1070	5	O8T3A7	O8T3A7 caenorhabdi
35	7.7	1111	5	O9XWD6	O9XWD6 caenorhabdi
36	7.7	1142	10	O8WQ06	O8WQ06 sorghum bic
37	7.7	1232	5	O9V7D0	O9V7D0 drosophila
38	7.7	1553	5	O95TR0	O95TR0 drosophila
39	7.7	1953	5	O9VKD0	O9VKD0 drosophila
40	7.7	2110	5	O9VRA6	O9VRA6 drosophila
41	7.7	2166	5	O9VK42	O9VK42 drosophila
42	7.7	2275	3	O93937	O93937 emericea
43	7.7	3199	3	O96UW4	O96UW4 ustilago ma
44	6.7	66	16	O97R39	O97R39 streptococc
45	6.7	88	5	O9W1U9	O9W1U9 drosophila

ALIGNMENTS

RESULT 1

ID	Q9LSX0	PRELIMINARY:	PRT:	131 AA.
AC	O9LSX0;			
DT	01-OCT-2000 (TREMREL. 15, Created)			
DT	01-OCT-2000 (TREMREL. 15, Last sequence update)			
DT	01-JUN-2002 (TREMREL. 21, Last annotation update)			
DE	Nicotiana EREBP-3 like.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=COLUMBIA;			
RX	MEDLINE=20181125; Pubmed=10718197;			
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,			
RA	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones."			
RL	DNA Res. 7:31-63(2000)			
DR	EMBL; AB025638; BAA97420.1;			
DR	HSSP; O80337; 26CC.			
DR	InterPro; IPR001471; TF_ERF.			
DR	Pfam; PF00847; AP2-domain; 1.			
DR	PRINTS; PR00367; ETHRSPELEMT.			
DR	ProDom; PD001423; TF_AP2; 1.			
DR	SMART; SM00380; AP2; 1.			
SQ	SEQUENCE 131 AA; 14343 MW; ABD1284C83167F61 CRC64;			

Query Match	8.7%;	Score 9;	DB 10;	Length 131;
Best Local Similarity	100.0%;	Pred. No. 0.26;		
Matches	9;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
Qy	53	SSASASASS	61	
Db	91	SSASASASS	99	

```

RESULT 2
Q9W404 PRELIMINARY; PRT: 406 AA.
ID Q9W404;
AC Q9W404;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CG3842 protein.
DE CG3842.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC EMBL: AE003437; AAF46156.1; -.
CC HSSP: P14061; 1FDU.
CC FlyBase: FBgn0029866; CG3842.
CC InterPro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short.1.
CC PRINTS: PR00080; SDRFAMILY.
CC Oxidoreductase.
SQ SSSEQUENCE 406 AA; 44889 MW; 85CCF99EC52A6D7D CRC64;

```

```

Query Match 8.7%; Score 9; DB 5; Length 406;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 53 SSASASASS 61
DB 379 SSASASASS 387

RESULT 3
Q9LT24 PRELIMINARY; PRT: 554 AA.
ID Q9LT24;
AC Q9LT24;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE G1AAD50054.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty pl and .7AC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL: AB025631; BAB01299.1; -.
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR004249; NPH3.
DR Pfam: PF03000; NPH3; 1.
DR SMART: SM00225; BTB; 1.
DR PROSITE: PS00097; BTB; 1.
SQ SEQUENCE 554 AA; 62991 MW; 93C37A098725927B CRC64;

Query Match 8.7%; Score 9; DB 10; Length 554;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASASS 61
DB 171 SSASASASS 179

RESULT 4
Q9DC19 PRELIMINARY; PRT: 588 AA.
ID Q9DC19;
AC Q9DC19;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Adult male lung cDNA, RIKEN full-length enriched library,
DE clone:1200007A08, full insert sequence.
GN MBDI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Hono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

```

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Wagner T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hashizaki Y.,
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
RL EMBL: AK004624; BAB23419.1; -.
DR MGD: MGI:1333811; Mbd1.
DR InterPro: IPR001739; Methyl-CpG_bind.
DR InterPro: IPR002857; Znf_CXXC.
DR Pfam: PF01429; MBD; 1.
DR Pfam: PF02008; zf-CXXC; 2.
DR SMART: SM00391; MBD; 1.
SQ SEQUENCE 588 AA; 64182 MW; 00AD7484FC204195 CRC64;
Query Match 8.7%; Score 9; DB 11; Length 588;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 SSASASASS 61
DB 135 SSASASASS 143
|||||
RESULT 5
Q24191 PRELIMINARY; PRT; 877 AA.
ID Q24191
AC Q24191
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Transcriptional repressor protein.
GN SCM OR CG9495.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96202481; PubMed=8625848;
RA Bornemann D., Miller E., Simon J.;
RT "The Drosophila Polycomb group gene Sex comb on midleg (Scm) encodes a
RL zinc finger protein with similarity to polyhomeotic protein.*;
RL Development 122:1621-1630(1996).
DR EMBL: U49793; AAB57632.1; -.
DR FlyBase: FBgn003334; Scm.
DR InterPro: IPR004092; Mbt.
DR InterPro: IPR001660; SAM.
DR Pfam: PF02820; mbt; 2.
DR Pfam: PF00536; SAM; 1.
DR SMART: SM00454; SAM; 1.
SQ SEQUENCE 877 AA; 93550 MW; 7859FD0C7B91589E CRC64;

Query Match 8.7%; Score 9; DB 5; Length 877;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASASS 61
DB 26 SSASASASS 34
|||||

RESULT 6
Q9VHA0

ID Q9VHA0 PRELIMINARY; PRT; 877 AA.
AC Q9VHA0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SCM protein.
GN SCM OR CG9495.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003683; AAF54419.1; -.
DR FlyBase: FBgn0003334; Scm.
DR InterPro: IPR004092; Mbt.
DR InterPro: IPR001660; SAM.
DR Pfam: PF02820; mbt; 2.
DR Pfam: PF00536; SAM; 1.
DR SMART: SM00454; SAM; 1.
SQ SEQUENCE 877 AA; 93520 MW; 7859FD0C6D76C6F1 CRC64;

Query Match 8.7%; Score 9; DB 5; Length 877;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASASS 61
DB 26 SSASASASS 34
|||||

RESULT 7

Q23984
ID Q23984 PRELIMINARY; PRT; 2186 AA.
AC Q23984;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TIGRIN precursor.
GN TIG OR CG11527.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95009506; PubMed=7924982;
RA Fogerty F.J., Fessler L.I., Bunch T.A., Yaron Y., Parker C.G.,
RA Nelson R.E., Brower D.L., Gullberg D., Fessler J.H.;
RT "Tiggrin, a novel Drosophila extracellular matrix protein that
RT functions as a ligand for Drosophila alpha PS2 beta PS integrins.";
RL Development 120:1747-1758(1994).
DR EMBL: U09506; AAA56998.1;
DR FlyBase: FBgn0011722; T1g.
DR InterPro: IPR002017; Spectrin.
KW Signal.
FT CHAIN 1 18 POTENTIAL.
FT SIGNAL 19 2186 TIGRIN.
SQ SEQUENCE 2186 AA; 257115 MW; AD24306A370E7D31 CRC64;
Query Match 8.7%; Score 9; DB 5; Length 2186;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 GSSASASAS 60
DB 2069 GSSASASAS 2077
IIIIIIIIII
GSSASASAS 2077
RESULT 8
Q9VMD9
ID Q9VMD9 PRELIMINARY; PRT; 2188 AA.
AC Q9VMD9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Tig protein.
GN TIG OR CG11527.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.G., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek G., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003613; AAF52380.2;
DR FlyBase: FBgn0011722; T1g.
DR InterPro: IPR002017; Spectrin.
SQ SEQUENCE 2188 AA; 257445 MW; FFD4715696C33E9E CRC64;
Query Match 8.7%; Score 9; DB 5; Length 2188;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 GSSASASAS 60
DB 2069 GSSASASAS 2077
IIIIIIIIII
GSSASASAS 2077
RESULT 9
Q9C754
ID Q9C754 PRELIMINARY; PRT; 97 AA.
AC Q9C754;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 10.6 kDa protein (At1g30260/F12p21_9).
GN F12p21_9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV COLUMBIA;
RA MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hulzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
RN [2]

```

RP SEQUENCE FROM N.A.
RA Chew R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC073506; AAG50559.1; -.
DR EMBL; AY057696; AAL15326.1; -.
KW Hypothetical protein.
SQ SEQUENCE 97 AA; 10587 MW; 8754363842C6864D CRC64;

Query Match 7.7%; Score 8; DB 10; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.2; 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 53 SSASASAS 60
Db 68 SSASASAS 75

RESULT 10
Q9XIA6 PRELIMINARY; PRT; 116 AA.
ID Q9XIA6
AC Q9XIA6
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Putative ribosomal protein.
GN Fl3p21.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Alcafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007504; AAD43160.1; -.
DR HSSP; P23828; Irip.
DR InterPro; IPR000266; Ribosomal_S17.
DR Pfam; PF00366; Ribosomal_S17; 1.
DR PRINTS; PR00973; RIBOSOMALS17.
DR ProDom; PD001295; Ribosomal_S17; 1.
KW Ribosomal protein.
SQ SEQUENCE 116 AA; 12560 MW; DA2C5D5F6D5F0E00 CRC64;

Query Match 7.7%; Score 8; DB 10; Length 116;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SASASASS 61
Db 91 SASASASS 98

RESULT 11
Q9V5U8 PRELIMINARY; PRT; 121 AA.
ID Q9V5U8
AC Q9V5U8
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG9080 protein.

```

```

GN CG9080.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balcer K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003826; AAF58699.1; -.
DR FlyBase; FBgn0033593; CG9080.
SQ SEQUENCE 121 AA; 11690 MW; 8047A2803CDA3338 CRC64;

Query Match 7.7%; Score 8; DB 5; Length 121;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SASASASS 61
Db 109 SASASASS 116

RESULT 12
Q9VJV7 PRELIMINARY; PRT; 156 AA.
ID Q9VJV7
AC Q9VJV7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG16866 protein.
GN CG16866.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003641; AAF5343.1;
 DR FlyBase; FBgn0032534; CG16866.
 SQ SEQUENCE 156 AA; 17040 MW; 30243C076B798CCF CRC64;

Query Match 7.7% Score 8; DB 5; Length 156;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 SSASASAS 60
 Db 56 SSASASAS 63
 |||||

RESULT 13
 Q9W1Q2 PRELIMINARY; PRT; 208 AA.
 AC Q9W1Q2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 GN CG9815 protein.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003641; AAF5343.1;
 DR FlyBase; FBgn0032534; CG16866.
 SQ SEQUENCE 156 AA; 17040 MW; 30243C076B798CCF CRC64;

RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003460; AAF47003.1;
 DR FlyBase; FBgn0034861; CG9815.
 SQ SEQUENCE 208 AA; 22081 MW; D3D52D1AD2ABA1CA CRC64;

Query Match 7.7% Score 8; DB 5; Length 208;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 SSASASAS 60
 Db 67 SSASASAS 74
 |||||

RESULT 14
 Q9KYH7 PRELIMINARY; PRT; 256 AA.
 AC Q9KYH7;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein SCO2695.
 GN SCO2695 OR SCC61A.16.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 RN NCBI_TaxID=1902;
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003460; AAF47003.1;
 DR FlyBase; FBgn0034861; CG9815.
 SQ SEQUENCE 208 AA; 22081 MW; D3D52D1AD2ABA1CA CRC64;

RA Huang C.-H., Kieser T., Larke L., Murphly L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL356595; CAB92261.1; --
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 27290 MW; 8AA505DD438C32AC CRC64; --

Query Match 7.7%; Score 8; DB 16; Length 256;
Best Local Similarity 100.0%; Pred. No. 4.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
Db 248 SSASASAS 255
|||||

RESULT 15

Q9N7U4

ID Q9N7U4 PRELIMINARY; PRT; 279 AA.

AC Q9N7U4;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE Possible putative aldehyde dehydrogenase (Fragment).

GN LM28.84.

OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI_TaxID=5664;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-FRIEDLIN;

RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL390935; CAC00843.1; --

FT NON_TER 1

FT NON_TER 279 279

SQ SEQUENCE 279 AA; 30259 MW; EB8380BB047F9FC5 CRC64;

Query Match

Best Local Similarity 100.0%; DB 5; Length 279;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 ASASASSL 62
Db 129 ASASASSL 136
|||||

Search completed: July 16, 2003, 12:52:46
Job time : 31 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 12:11:40 ; Search time 80 Seconds
(without alignments)
5965.064 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 375

Sequence: 1 aatgagcaacgtgtcaaca.....cctctctgtatcaatgcttt 1158

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2.1/USPTO.spool/FASTAN -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0
-DB=SPTRMBL_21 -QSWT=usfan -TRANS=human40.cdi -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -MAXLEN=2000000000
-USER=US09814661_@CGN_1.139 @runat 16072003.115331.15205 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	11	2.9	733	13	Q9PTL8	Q9ptl8 xenopus lae

2	11	2.9	2206	5	096205	096205 plasmodium
3	10	2.7	85	11	Q61402	Q61402 mus musculus
4	10	2.7	89	12	Q8Q003	Q8q003 camelopard vi
5	10	2.7	97	12	Q89696	Q89696 variola vir
6	10	2.7	97	16	Q9JYU5	Q9jyu5 neisseria m
7	10	2.7	99	12	Q8Q004	Q8q004 camelopard vi
8	10	2.7	102	11	Q9D1W2	Q9d1w2 mus musculus
9	10	2.7	112	3	Q03884	Q03884 saccharomyc
10	10	2.7	113	10	Q9LP60	Q9lp60 arabidopsis
11	10	2.7	269	5	096188	096188 plasmodium
12	10	2.7	321	5	Q19051	Q19051 caenorhabdi
13	10	2.7	321	5	Q19051	Q19051 caenorhabdi
14	10	2.7	322	5	P90550	P90550 leishmania
15	10	2.7	400	5	Q95YR4	Q95yr4 leishmania
16	10	2.7	415	4	Q8TCB8	Q8tc88 homo sapien
17	10	2.7	605	10	Q9SH28	Q9sh28 arabidopsis
18	10	2.7	1129	5	Q93342	Q93342 caenorhabdi
19	10	2.7	1161	5	Q9TY17	Q9ty17 plasmodium
20	10	2.7	1266	5	Q95258	Q95258 plasmodium
21	9	2.4	35	10	Q9LQ64	Q9lq64 arabidopsis
22	9	2.4	58	12	Q8V5S2	Q8v5s2 helicoverpa
23	9	2.4	86	12	Q89191	Q89191 variola vir
24	9	2.4	91	10	Q9FIY3	Q9fiy3 arabidopsis
25	9	2.4	101	5	Q9N8G1	Q9n8g1 trypanosoma
26	9	2.4	113	10	Q9LP60	Q9lp60 arabidopsis
27	9	2.4	125	3	P87268	P87268 saccharomyc
28	9	2.4	125	5	Q9N8G0	Q9n8g0 trypanosoma
29	9	2.4	129	5	Q95RT0	Q95rt0 drosophila
30	9	2.4	131	10	Q9LSX0	Q9lsx0 arabidopsis
31	9	2.4	307	5	Q96261	Q96261 plasmodium
32	9	2.4	406	5	Q9W404	Q9w404 drosophila
33	9	2.4	426	10	Q04662	Q04662 arabidopsis
34	9	2.4	554	10	Q9LT24	Q9lt24 arabidopsis
35	9	2.4	588	11	Q9DC19	Q9dc19 mus musculus
36	9	2.4	682	10	Q9LMV4	Q9lmv4 arabidopsis
37	9	2.4	877	5	Q24191	Q24191 drosophila
38	9	2.4	877	5	Q9VHA0	Q9vha0 drosophila
39	9	2.4	1221	5	Q24079	Q24079 drosophila
40	9	2.4	1866	5	Q23984	Q23984 drosophila
41	9	2.4	2188	5	Q9VMD9	Q9vmd9 drosophila
42	9	2.4	2206	5	Q96205	Q96205 plasmodium
43	9	2.4	2391	5	Q27732	Q27732 plasmodium
44	9	2.4	35	10	Q9LQ64	Q9lq64 arabidopsis
45	8	2.2	35	10	Q9LQ64	Q9lq64 arabidopsis

ALIGNMENTS

RESULT 1
Q9PTL8 PRELIMINARY; PRT; 733 AA.
ID Q9PTL8
AC Q9PTL8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F-box protein 26 (Fragment).
GN FBX26.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Regan Reimann J.D., Duong Q.V., Jackson P.K.;
RT "Identification of novel F-box proteins in Xenopus laevis.";
RL Curr. Biol. 9:0-0(2000).
DR EMBL; AF176353; AAF14554.1;
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
FT NON_TER 1
SQ SEQUENCE 733 AA; 82371 MW; 1A982E8DA475634E CRC64;

Alignment Scores:

Pred. No.: 0.0195 Length: 733
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.93% Indels: 0
 DB: 13 Gaps: 0

US-09-814-661A-1 (1-1158) x Q9PPL8 (1-733)

QY 1010 TTTATACATATATATATATATATATATATAT 1042

Db 663 PheIleTyrlleTyrlleTyrlleTyrlleTyrl 673

RESULT 2

Q96205 PRELIMINARY; PRT; 2206 AA.

AC Q96205; (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Hypothetical 264.1 kDa protein.

CN PF80560W.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99021743; PubMed=9804551;

RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,

RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,

RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,

RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,

RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;

RT "Chromosome 2 sequence of the human malaria parasite Plasmodium

RT falciparum";

RL Science 282:1126-1132(1998).

DR EMBL; AE001403; AAC71901.1; -

KW Hypothetical protein.

SQ SEQUENCE 2206 AA; 264100 MW; 902C190DCE5B7901 CRC64;

Alignment Scores:

Pred. No.: 0.0174 Length: 2206
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.93% Indels: 0
 DB: 5 Gaps: 0

US-09-814-661A-1 (1-1158) x Q96205 (1-2206)

QY 1013 ATATACATATATATATATATATATATATATGTC 1045

Db 2196 IleTyrlleTyrlleTyrlleTyrlleTyrlVal 2206

RESULT 3

Q61402 PRELIMINARY; PRT; 85 AA.

AC Q61402; (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Gcapi protein (Fragment).

DE Gcapi protein (Fragment).

GN GCAP8 OR GCAP1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CEREBELLUM;

RA Kambouris M.;

RL Thesis (1993), Unknown Institution.

RN SEQUENCE FROM N.A.
 RP TISSUE=CEREBELLUM;
 RX MEDLINE=95107035; PubMed=7808217;
 RA Kambouris M., Sangameswaran L., Triarhou L.C., Kozak C.A.,
 RA Dlouhy S.R., Ghetti B., Hodes M.E.;
 RT "Molecular characterization of a novel cDNA from murine cerebellum,
 RT developmental expression, and distribution in brain.";
 RL Brain Res. Mol. Brain Res. 25:192-199(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CEREBELLUM;
 RX MEDLINE=95107034; PubMed=7808216;
 RA Kambouris M., Triarhou L.C., Dlouhy S.R., Sangameswaran L., Luo F.,
 RA Ghetti B., Hodes M.E.;
 RT "Novel cDNA clones obtained by antibody screening of a mouse
 RT cerebellar cDNA expression library.";
 RL Brain Res. Mol. Brain Res. 25:183-191(1994).
 DR EMBL; L10908; AAA68426.1; -
 DR MGD; MGI:95672; Gcap8.
 FT NON_TER 1
 SQ SEQUENCE 85 AA; 10115 MW; 0E4710985D022083 CRC64;

Alignment Scores:

Pred. No.: 0.299 Length: 85
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 11 Gaps: 0

US-09-814-661A-1 (1-1158) x Q61402 (1-85)

Qy 1013 ATATACATATATATATATATATATATATATAT 1042

Db 55 IleTyrlleTyrlleTyrlleTyrlleTyrl 64

RESULT 4

Q8QQ03 PRELIMINARY; PRT; 89 AA.

AC Q8QQ03; (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE CML190.5BL

GN CML190.5BL

OS Canelpox virus (strain CP-1).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OC Orthopoxvirus.

OX NCBI_TaxID=28873;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CMS;

RX PubMed=11907336;

RA Gubser C., Smith G.L.;

RT "The sequence of canelpox virus shows it is most closely related to

RT variola virus, the cause of smallpox.";

RL J. Gen. Virol. 83:855-872(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CMS;

RA Gubser C., Smith G.L.;

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY009089; AAG37698.1; -

SQ SEQUENCE 89 AA; 10585 MW; 739D5689840FF25A CRC64;

Alignment Scores:

Pred. No.: 0.298 Length: 89
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 12 Gaps: 0

US-09-814-661A-1 (1-1158) x Q8QQ03 (1-89)

QY 1013 ATATACATATATATATATATATATAT 1042

Db 6 lletyrlletyrlletyrlletyrlletyrl 15

RESULT 5

Q89696 PRELIMINARY; PRT; 97 AA.
 AC Q89696;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ORF15L.
 GN D5L.
 OS Variola virus, and
 OS variola minor virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10255, 53258;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Variola virus; STRAIN=GARCIA-1966;
 RA Shchelkunov S.N., Totmenin A.V., Resenchuk S.M., Blinov V.M.,
 RA Sandakhchiev L.S.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Variola virus; STRAIN=GARCIA-1966;
 RA Shchelkunov S.N.;
 RL Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Variola virus; STRAIN=GARCIA-1966;
 RX MEDLINE=93202281; PubMed=8384129;
 RA Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
 RA "Genetics of variola and vaccinia viruses necessary to overcome the host
 RT protective mechanisms";
 RL FEBS Lett. 319:80-83(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Variola virus; STRAIN=GARCIA-1966;
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RL "The complete DNA sequence of vaccinia virus.";
 RL Virology 179:247-266(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Variola virus; STRAIN=GARCIA-1966;
 RA Shchelkunov S.N.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Variola virus; STRAIN=GARCIA-1966;
 RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
 RA Totmenin A.V., Shchelkunov S.N., Esposito J.J.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES-variola minor virus; STRAIN=GARCIA-1966;
 RA Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
 RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
 RA Esposito J.J., Sosnovtsev S.;
 RL "Analysis of the complete coding sequence of DNA of alastrim variola
 RT minor virus strain Garcia-1966.";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; X72086; CAA50966.1; -;
 DR EMBL; U18339; AAA69395.1; -;
 DR EMBL; Y16780; CAB54786.1; -;
 SQ SEQUENCE 97 AA; 11783 MW; A3AD562E1B756A49 CRC64;

Alignment Scores:

Pred. No.: 0.295 Length: 97
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 12 Gaps: 0

US-09-814-661A-1 (1-1158) x Q89696 (1-97)

QY 1013 ATATACATATATATATATATATATATAT 1042

Db 6 lletyrlletyrlletyrlletyrlletyrl 15

RESULT 6

Q89696 PRELIMINARY; PRT; 97 AA.
 ID Q89696;
 AC Q89696;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical protein NMA0677.
 GN NMA0677.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=2022556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrrell B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria
 RT meningitidis Z2491";
 RL Nature 404:502-506(2000).
 DR EMBL; AL162753; CAB83964.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 97 AA; 10612 MW; 62046849BF32E4E9 CRC64;

Alignment Scores:

Pred. No.: 0.295 Length: 97
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 16 Gaps: 0

US-09-814-661A-1 (1-1158) x Q89696 (1-97)

QY 949 ACTCTCGTCTACTATTGTGTCTATCTCG 978

Db 3 ThrLeuValLeuLeuLeuSerPheSer 12

RESULT 7

Q8QQ04 PRELIMINARY; PRT; 99 AA.
 ID Q8QQ04;
 AC Q8QQ04;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE CMP190R.
 GN CMP190R.
 OS Camelpox virus (strain Cp-1).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=28873;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMS;

RN SEQUENCE FROM N.A.
 RP Ecker J.R.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
 Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC020889; AAF79706.1; -;
 DR InterPro; IPR000477; RVTse.
 DR Pfam; PF00078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 113 AA; 13053 MW; 0958B4933D3F7E46 CRC64;

Alignment Scores:
 Pred. No.: 0.291 Length: 113
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 10 Gaps: 0

US-09-814-661A-1 (1-1158) x O9LP60 (1-113)

QY 1016 TACATATATATATATATATATATATATGTC 1045
 Db 82 TyrIleTyrIleTyrIleTyrVal 91

RESULT 11
 O96188
 ID O96188 PRELIMINARY; PRT; 269 AA.
 AC O96188;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE Predicted multiple transmembrane domain protein.
 GN PF00475C.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99021743; PubMed=9804551;
 RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
 Koonin E.V., Shalloom S., Mason T., Yu K., Fujii C., Pederson J.,
 Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perteau M.,
 Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
 Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
 falciparum.";
 RL Science 282:1126-1132(1998).
 DR EMBL; AE001397; AAC71884.1; -;
 KW Transmembrane.
 SQ SEQUENCE 269 AA; 32703 MW; F6995A40862E28C9 CRC64;

Alignment Scores:
 Pred. No.: 0.265 Length: 269
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.70% Indels: 0
 DB: 5 Gaps: 0

US-09-814-661A-1 (1-1158) x O96188 (1-269)
 QY 1041 TATATATATATATATATATATATATATA 1012
 Db 147 TyrIleTyrIleTyrIleTyrMetTyrIle 156

RESULT 12
 Q19051
 ID Q19051 PRELIMINARY; PRT; 321 AA.
 AC Q19051;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE E03G2.1 protein.
 GN E03G2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditida; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.A.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]

Alignment Scores:
 Pred. No.: 0.261 Length: 321
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 5 Gaps: 0

US-09-814-661A-1 (1-1158) x Q19051 (1-321)

QY 1016 TACATATATATATATATATATATATGTC 1045
 Db 85 TyrIleTyrIleTyrIleTyrVal 94

RESULT 13
 Q19051
 ID Q19051 PRELIMINARY; PRT; 321 AA.
 AC Q19051;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE E03G2.1 protein.
 GN E03G2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditida; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.A.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]

Alignment Scores:
 Pred. No.: 0.261 Length: 321
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.67% Indels: 0
 DB: 5 Gaps: 0

US-09-814-661A-1 (1-1158) x Q19051 (1-321)

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 15:13:54 ; Search time 1643 Seconds
(without alignments)
11414.713 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 1158

Sequence: 1 aatgagcaaccgtgtcaaca.....cctctctgatcaatgcttt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	351	30.3	589	17	AQ873494
c 2	350	30.2	526	17	AQ873357
c 3	271	23.4	573	17	AQ872886
c 4	254	21.9	466	17	AQ501881
c 5	228	19.7	535	17	AQ502431
c 6	221	19.1	562	17	AQ500866

219	18.9	499	17	AQ873058	AQ873058	V58E5	mtf
206	17.8	784	17	AQ875860	AQ875860	V130C9	mtf
144	12.4	590	17	AQ873374	AQ873374	V64G3	mtf
139	12.0	543	17	AQ500106	AQ500106	V30P8	mtf
126	10.9	796	17	AQ875934	AQ875934	V131C10	m
76	6.6	543	17	AQ501433	AQ501433	V22C5	mtf
67	5.8	530	17	AQ503366	AQ503366	V56E10	mtf
46	4.0	525	17	AQ502957	AQ502957	V47B7	mtf
38	3.3	619	17	AQ500377	AQ500377	V40F1	mtf
35	3.0	475	14	B0749538	B0749538	SnEP4a65	
34	2.9	286	10	BB024373	BB024373	BB024373	
33	2.9	420	9	AI099986	AI099986	34139	Lam
33	2.8	191	17	AZ762398	AZ762398	1M0557M23	
33	2.8	317	13	BI670816	BI670816	PfEST0aa0	
33	2.8	355	17	AZ518177	AZ518177	RPCI-11-2	
33	2.8	376	17	AZ055998	AZ055998	RPCI-23-4	
33	2.8	376	17	AZ055998	AZ055998	RPCI-23-4	
33	2.8	394	17	AQ657248	AQ657248	Sheared D	
33	2.8	410	17	AQ010422	AQ010422	HS_2172_B	
26	2.8	423	12	BF325001	BF325001	su18b02.y	
33	2.8	450	17	DR4J6T	DR4J6T	Danio rer	
28	2.8	468	9	AI803880	AI803880	tp30f02.x	
29	2.8	475	13	BI814006	BI814006	PfEST0aa2	
30	2.8	483	10	AW832093	AW832093	sm19g09.y	
33	2.8	497	17	AZ605889	AZ605889	1M0427C01	
32	2.8	522	13	BI891319	BI891319	2F637-3-0	
33	2.8	527	17	AZ778393	AZ778393	2M0013102	
33	2.8	529	17	AZ294884	AZ294884	RPCI-23-1	
33	2.8	529	17	AZ294884	AZ294884	RPCI-23-1	
33	2.8	537	17	AZ497263	AZ497263	1M0334J02	
33	2.8	547	17	AQ873740	AQ873740	V74E5	mtf
33	2.8	550	17	AQ419130	AQ419130	RPCI-11-1	
33	2.8	554	17	AZ752028	AZ752028	RPCI-24-1	
33	2.8	571	13	BM084974	BM084974	sa130e07	
33	2.8	576	17	BM435516	BM435516	BOG9259F	
33	2.8	579	14	BQ452669	BQ452669	sao89h11	
33	2.8	596	17	AZ552500	AZ552500	RPCI-23-2	
33	2.8	596	17	AZ552500	AZ552500	RPCI-23-2	
33	2.8	606	17	AZ987345	AZ987345	2M0269F20	

ALIGNMENTS

RESULT 1
LOCUS AQ873494/C
DEFINITION V66G5 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.
ACCESSION AQ873494
VERSION AQ873494.1 GI:6285738
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 589)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symonlatas, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9549
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCTTCTTCTTTTGAAGTAC

Class: transposon-tagged.
 Location/Qualifiers
 1..589
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS56-Sal; A yeast genomic DNA library was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 162 a 125 c 129 g 173 t
 ORIGIN
 Query Match 30.3%; Score 351; DB 17; Length 589;
 Best Local Similarity 99.8%; Pred. No. 4e-147;
 Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 38 TACTTATTCCTCCCAAGGATCAGTTCCTCTCTGCAACATCATTCGCGTGAACGTCGCGGC 97
 Db |||||||
 QY 469 TACTTATTCCTCCCAAGGATCAGTTCCTCTCTGCAACATCATTCGCGTGAACGTCGCGGC 410
 Db |||||||
 QY 98 GGTCTTTCTGACATTGGTAAGATACTTCCAACTAAGAGCATGCTCTCTTTTTTTTGT 157
 Db |||||||
 QY 409 GGTCTTTCTGACATTGGTAAGATACTTCCAACTAAGAGCATGCTCTCTTTTTTTTGT 350
 Db |||||||
 QY 158 AGGCCAATGATAGGAAGAACAATAGATTATAATACGTCAGATATATAGTAGATATGTTT 217
 Db |||||||
 QY 349 AGGCCAATGATAGGAAGAACAATAGATTATAATACGTCAGATATATAGTAGATATGTTT 290
 Db |||||||
 QY 218 TTATCTTTAGACCTCGTACATAGGAATAATGACCTTTTTTTTGGCCAAACATTTGAAT 277
 Db |||||||
 QY 289 TTATCTTTAGACCTCGTACATAGGAATAATGACCTTTTTTTTGGCCAAACATTTGAAT 230
 Db |||||||
 QY 278 TTTTCTTTTGTACCTCGCGCTGAGCCCAACGGGCTCCACTACCGCGCGGTCGCCATT 337
 Db |||||||
 QY 229 TTTTCTTTTGTACCTCGCGCTGAGCCCAACGGGCTCCACTACCGCGCGGTCGCCATT 170
 Db |||||||
 QY 338 TTGGGAAGTCATCCGTCGCAAAAGGAATAGCCATACATATCCTTTACTGTTTGGAAAC 397
 Db |||||||
 QY 169 TTGGGAAGTCATCCGTCGCAAAAGGAATAGCCATACATATCCTTTACTGTTTGGAAAC 110
 Db |||||||
 QY 398 ATCGCCCGTTTCGCCCGATTCCGCTCAGCGGGTATAAAAG 439
 Db |||||||
 QY 109 ATCGCCCGTTTCGCCCGATTCCGCTCAGCGGGTATAAAAG 68
 Db |||||||

RESULT 2
 A0873357/c
 LOCUS
 DEFINITION V6402 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.
 ACCESSION A0873357
 VERSION A0873357.1 GI:6285601
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 526)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University

P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mTn-3xHA/lacZ insertion.
 Seq primer: GCCCTCTTCTTTGGAAAGTAC
 Class: transposon-tagged.

FEATURES
 Location/Qualifiers
 source 1..526
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS56-Sal; A yeast genomic DNA library was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 159 a 109 c 116 g 141 t 1 others
 ORIGIN
 Query Match 30.2%; Score 350; DB 17; Length 526;
 Best Local Similarity 100.0%; Pred. No. 1.2e-146;
 Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 120 ATACTTCAACTAAGAGCATGCTCTCTCTTTTTTTTGTAGGCCAATGATAGGAAGAACA 179
 Db |||||||
 QY 428 ATACTTCAACTAAGAGCATGCTCTCTCTTTTTTTTGTAGGCCAATGATAGGAAGAACA 369
 Db |||||||
 QY 180 ATAGATTATAAATACGTCAGATATAGTAGATATGTTTATGTTAGACCTCGTACATA 239
 Db |||||||
 QY 368 ATAGATTATAAATACGTCAGATATAGTAGATATGTTTATGTTAGACCTCGTACATA 309
 Db |||||||
 QY 240 GAATAAATGACGTTTTTTTTTGGCCAAACATTTGAAATTTTTTTTTTTGTTACCTCGCGTG 299
 Db |||||||
 QY 308 GAATAAATGACGTTTTTTTTTGGCCAAACATTTGAAATTTTTTTTTTTGTTACCTCGCGTG 249
 Db |||||||
 QY 300 AGCCCAACGGGCTCCACTACCGCGCGGTCGCCATTTCGGAGTGCATCCGTCGCCA 359
 Db |||||||
 QY 248 AGCCCAACGGGCTCCACTACCGCGCGGTCGCCATTTCGGAGTGCATCCGTCGCCA 189
 Db |||||||
 QY 360 AAGAAATAGCCATAACATATCGTTACTGTTTTTGGAAAGTGCATCCGTCGCCA 419
 Db |||||||
 QY 188 AAGAAATAGCCATAACATATCGTTACTGTTTGGAAAGTGCATCCGTCGCCA 129
 Db |||||||
 QY 420 GCCTCAGCGGGTATAAAAGAGATCTTTTTTTTCTCGCTGCTCCCTCC 469
 Db |||||||
 QY 128 GCCTCAGCGGGTATAAAAGAGATCTTTTTTTTCTCGCTGCTCCCTCC 79
 Db |||||||

RESULT 3
 A0872886
 LOCUS
 DEFINITION V549 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.
 ACCESSION A0872886
 VERSION A0872886.1 GI:6285130
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 573)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mtn-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTTTGGAGTAC
Class: transposon-tagged.
Location/Qualifiers
1. .573

FEATURES
source

/organism="Saccharomyces cerevisiae"
/strain="AB972 - trp1 r(0) (5288C background)"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 119 a 131 c 94 g 229 t

ORIGIN

Query Match 23.4%; Score 271; DB 17; Length 573;
Best Local Similarity 100.0%; Pred. No. 3.8e-111;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 GAATACCTCACTAGACATGCTCTCTTTTGTAGCCCAATGATGAAGAA 177
DB 68 GAATACCTCACTAGACATGCTCTCTTTTGTAGCCCAATGATGAAGAA 127
QY 178 CAATAGATTATAATACGTGAGATATAGTAGATGTTTATGTTTACCTGCTGACA 237
DB 128 CAATAGATTATAATACGTGAGATATAGTAGATGTTTATGTTTACCTGCTGACA 187
QY 238 TAGGAATAATGACGTTTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGCGC 297
DB 188 TAGGAATAATGACGTTTTTTTTTGGCCCAACATTTGAAATTTTTTTTGTACCTCGCGC 247
QY 298 TAGAGCCCAACGGCTCCACTACCGCGCGTCCGATTTGGGAAGTCATCCGTCCTCA 357
DB 248 TAGAGCCCAACGGCTCCACTACCGCGCGTCCGATTTGGGAAGTCATCCGTCCTCA 307
QY 358 AAAAGGAATAGCCATAACATATCGTTACTG 388
DB 308 AAAAGGAATAGCCATAACATATCGTTACTG 338

RESULT 4
AQ501881/c

LOCUS V13D1 mtn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', DNA sequence.
ACCESSION AQ501881
VERSION AQ501881.1 GI:4707531
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 466)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161

P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mtn-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTTTGGAGTAC
Class: transposon-tagged.
Location/Qualifiers
1. .466

FEATURES
source

/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 131 a 92 c 120 g 119 t

ORIGIN

Query Match 21.9%; Score 254; DB 17; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.8e-103;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 497 GTGATCTTACGGTCTCACTAACCTCTTCAACTGCTCAATAATTTCCGCTATGCAAAA 556
DB 319 GTGATCTTACGGTCTCACTAACCTCTTCAACTGCTCAATAATTTCCGCTATGCAAAA 260
QY 557 TTCCCAAGACTACTTTTACGCTCAAAATGCTGCCAACACAAACAGCCCTTCCACATT 616
DB 259 TTCCCAAGACTACTTTTACGCTCAAAATGCTGCCAACACAAACAGCCCTTCCACATT 200
QY 617 GGTACCGTACCATCGCGGAATTTAGAAGGTGCGCTTGGCCACCTATGCTGAGGTTC 676
DB 199 GGTACCGTACCATCGCGGAATTTAGAAGGTGCGCTTGGCCACCTATGCTGAGGTTC 140
QY 677 TATGTTGTCTACTCAAAACTCCATGGCAGCTCCGCTTCTGCTCCGCTTCTTCATTAGA 736
DB 139 TATGTTGTCTACTCAAAACTCCATGGCAGCTCCGCTTCTGCTCCGCTTCTTCATTAGA 80
QY 737 AATGTGGGAAAAGG 750
DB 79 AATGTGGGAAAAGG 66

RESULT 5
AQ502431/c

LOCUS V40D6 mtn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', DNA sequence.
ACCESSION AQ502431
VERSION AQ502431.1 GI:4708081
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 535)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161

```

Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source            1..562
                        /organism="Saccharomyces cerevisiae"
                        /db_xref="taxon:4932"
                        /clone_lib="mtfn-3xHA/lacZ Insertion Library"
                        /lab_host="E. coli"
                        /note="vector: PHSS6-Sal; A yeast genomic DNA library
                                (lacking mitochondrial DNA) was prepared in PHSS6-Sal;
                                genomic DNA was size-fractionated (DNA of roughly 2-3 kb
                                in length) prior to cloning. This library was
                                subsequently mutagenized with a mtn-3xHA/lacZ
                                minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT            195 a   87 c  127 g  153 t

```

[illegible]

A0873058 499 bp DNA linear GSS 08-NOV-1999
 LOCUS V58E5 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomycetes
 DEFINITION cerevisiae genomic 5', DNA sequence.
 ACCESSION A0873058
 VERSION A0873058.1 GI:6285302
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomycetes cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 499)
 Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
 desEvisages, S. A., Cheung, K. -H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umansky, L., Heldtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 Unpublished (1999)
 CONTACT: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 seq of mTn-3xHA/lacZ Insertion.
 Seq primer: GGCTTCCTTCCTTTGGAGTAC
 Class: transposon-tagged.
 Location/Qualifiers
 1. .499
 /organism="Saccharomycetes cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library, strain AB972"
 FEATURES
 source

/lab_host="E. coli"
/note="Vector: pHSS6-Sal; A yeast genomic DNA library was prepared in pHSS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

```
BASE COUNT      141 a   112 c   107 g   138 t      1 others
ORIGIN

Query Match      18.9%; Score 219; DB 17; Length 499;
Best Local Similarity 99.6%; Pred. No. 9.7e-88;
Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 175 GAACAATAGATATATAATACGTCAGAAATATAGTAGATATGTTTATGTTTAGACCTCGT 234
    |||||||
Db 383 GAACAATAGATATATAATACGTCAGAAATATAGTAGATATGTTTATGTTTAGACCTCGT 324
    |||||||

QY 235 ACATAGGAATATGACGTTTTTTTTTGGCCAAACATTTGAAATTTTTTTTGTACCTCG 294
    |||||||
Db 323 ACATAGGAATATGACGTTTTTTTTTGGCCAAACATTTGAAATTTTTTTTGTACCTCG 264
    |||||||

QY 295 CGCTAGCCCAACCGGCTCCACTACCGCGCGTCCCATTTTGGGAAGTCATCCGTC 354
    |||||||
Db 263 CGATAGCCCAACCGGCTCCACTACCGCGCGTCCCATTTTGGGAAGTCATCCGTC 204
    |||||||

QY 355 CCAAAAGGAATAGCCATACATATCGTTACTGTTTTTGGAAACATCGCCGTTTCGCCCG 414
    |||||||
Db 203 CCAAAAGGAATAGCCATACATATCGTTACTGTTTTTGGAAACATCGCCGTTTCGCCCG 144
    |||||||

QY 415 ATTCCGCTCAGCGGGTATATAAAGAGATC 444
    |||||||
Db 143 ATTCCGCTCAGCGGGTATATAAAGAGATC 114
    |||||||
```

```
RESULT 8
A0875860/c
LOCUS      A0875860      784 bp      DNA      linear      GSS 08-NOV-1999
DEFINITION V130C9 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces cerevisiae genomic 5', DNA sequence.
```

```
ACCESSION  A0875860
VERSION     A0875860.1  GI:6288104
KEYWORDS    GSS.
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
```

```
REFERENCE   1 (bases 1 to 784)
AUTHORS     Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatidis,D., Jansen,R.,
            Umansky,L., Heidman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
            Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
            Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
            Gene Disruption
            Unpublished (1999)
```

```
JOURNAL     Contact: Kumar A
COMMENT      Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
            Yale University
            P.O. Box 208103, New Haven, CT 06520-8103, USA
            Tel: 203 432 9949
            Fax: 203 432 6161
            Email: anuj.kumar@yale.edu
            te of mTn-3xHA/lacZ insertion.
            Seq primer: GGCCTTCTTCTTTGGAGTAC
            Class: transposon-tagged.
            Location/Qualifiers
                1..784
                    /organism="Saccharomyces cerevisiae"
                    /strain="Y2278 - S288C background, cir(0) rho(0)"
                    /db_xref="taxon:4932"
                    /clone_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"
                    /lab_host="E. coli"
                    /note="Vector: pHSS6-Sal; A yeast genomic DNA library
```

```
FEATURES
    source
        1..784
            /organism="Saccharomyces cerevisiae"
            /strain="Y2278 - S288C background, cir(0) rho(0)"
            /db_xref="taxon:4932"
            /clone_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"
            /lab_host="E. coli"
            /note="Vector: pHSS6-Sal; A yeast genomic DNA library
```

without 2 micron or mitochondrial DNA was prepared in pHSS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

```
BASE COUNT      236 a   147 c   202 g   199 t
ORIGIN

Query Match      17.8%; Score 206; DB 17; Length 784;
Best Local Similarity 99.7%; Pred. No. 5.9e-82;
Matches 326; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 333 CCATTTTGGGAGTCATCCGTCCTCCAAAAGGAAATAGGCATATACATATCGTTTACTGTTT 392
    |||||||
Db 464 CCATTTTGGGAGTCATCCGTCCTCCAAAAGGAAATAGGCATATACATATCGTTTACTGTTT 405
    |||||||

QY 393 GGAACATCGCCGGTTTCGCCCGATTCGCCCTCAGCGGGTATATAAAGAGATCTTTTTTTT 452
    |||||||
Db 404 GGAACATCGCCGGTTTCGCCCGATTCGCCCTCAGCGGGTATATAAAGAGATCTTTTTTTT 345
    |||||||

QY 453 TCCTGCGTGTCCCTT -CCATTTTAAATGCTTATCTGCTCCTTTGTGATCTTACGGTCT 511
    |||||||
Db 344 TCCTGCGTGTCCCTTCCCATTTTAAATGCTTATCTGCTCCTTTGTGATCTTACGGTCT 285
    |||||||

QY 512 CACTAACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATTTCCCAAGACTACTT 571
    |||||||
Db 284 CACTAACCTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATTTCCCAAGACTACTT 225
    |||||||

QY 572 TTACGCTCAAAATCGCTGCCAACACACACAGCCCTTCCCATTCGCGTACCGTGACCAT 631
    |||||||
Db 224 TTACGCTCAAAATCGCTGCCAACACACACAGCCCTTCCCATTCGCGTACCGTGACCAT 165
    |||||||

QY 632 GCGGAATTTAGAAGGTGCTTTGCC 658
    |||||||
Db 164 GCGGAATTTAGAAGGTGCTTTGCC 138
    |||||||
```

```
RESULT 9
A0873374/c
LOCUS      A0873374      590 bp      DNA      linear      GSS 08-NOV-1999
DEFINITION V64G3 mTn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.
```

```
ACCESSION  A0873374
VERSION     A0873374.1  GI:6285618
KEYWORDS    GSS.
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
```

```
REFERENCE   1 (bases 1 to 590)
AUTHORS     Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatidis,D., Jansen,R.,
            Umansky,L., Heidman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
            Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
            Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
            Gene Disruption
            Unpublished (1999)
```

```
JOURNAL     Contact: Kumar A
COMMENT      Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
            Yale University
            P.O. Box 208103, New Haven, CT 06520-8103, USA
            Tel: 203 432 9949
            Fax: 203 432 6161
            Email: anuj.kumar@yale.edu
            te of mTn-3xHA/lacZ insertion.
            Seq primer: GGCCTTCTTCTTTGGAGTAC
            Class: transposon-tagged.
            Location/Qualifiers
                1..590
                    /organism="Saccharomyces cerevisiae"
                    /strain="AB972 - trpl r(0) (S288C background)"
                    /db_xref="taxon:4932"
```

```
FEATURES
    source
        1..590
            /organism="Saccharomyces cerevisiae"
            /strain="AB972 - trpl r(0) (S288C background)"
            /db_xref="taxon:4932"
```

/clone_lib="mTn-3xHA/lacZ Insertion Library, strain AB972"
/lab_host="E. coli"
/note="vector: pHS56-Sal; A yeast genomic DNA library was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 153 a 129 c 120 g 188 t
ORIGIN

Query Match 12.4%; Score 144; DB 17; Length 590;
Best Local Similarity 99.5%; Pred. No. 4.9e-54;
Matches 194; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 102 TTTCTGACATTTGGTAAGAACTACTCCCAACTAAGACGCTCTCTCTTTTCTAGGC 161
DB 401 TTTCTGACATTTGGTAAGAACTACTCCCAACTAAGACGCTCTCTCTTTTCTAGGC 342
QY 162 CAATGATAGGAAGAACAATAGATTATATAATACGTCAGAATATAGTAGATATGTTTTAT 221
DB 341 CAATGATAGGAAGAACAATAGATTATATAATACGTCAGAATATAGTAGATATGTTTTAT 282
QY 222 GTTTAGACCTCGTACATAGGAATATTCACGCTTTTTTTTGGCCAAACATTTGAAATTTT 281
DB 281 GTTTAGACCTCGTACATAGGAATATTCACGCTTTTTTTTGGCCAAACATTTGAAATTTT 222
QY 282 TTTTGTACCTCGCG 296
DB 221 TTTTGTACCTCGCG 207

RESULT 10
A0500106/c
LOCUS V30F8 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
DEFINITION genomic 5', DNA sequence.
ACCESSION A0500106
VERSION A0500106.1 GI:4704928
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 543)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., desEtages, S. A., Cheung, K. H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCTCTCTTTCTTTGGAGATAC
Class: transposon-tagged.
Location/Qualifiers 1..543

FEATURES
source
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/note="vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ

minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 172 a 78 c 118 g 175 t
ORIGIN

Query Match 12.0%; Score 139; DB 17; Length 543;
Best Local Similarity 99.5%; Pred. No. 9e-52;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 969 GTCATTTCTCGAAACATTTCTCCCGTCGATTTTCCCTTTCCCTTATATACATATATAT 1028
DB 434 GTCATTTCTCGAAACATTTCTCCCGTCGATTTTCCCTTTATATACATATATATAT 375
QY 1029 ATATATATATATATGCTCTCTTCTACGATTTTTTGTATTTCTGTGCTTTTATCAAGATAG 1088
DB 374 ATATATATATATATGCTCTCTTCTAGGATTTTTCGATTTTCTGTGCTTTTATCAAGATAG 315
QY 1089 TCTATATACGTTTTGATACAGCTAGATATCGCTAGCGCAACATTTGCCCTCTCTTTGA 1148
DB 314 TCTATATACGTTTTGATACAGCTAGATATCGCTAGCGCAACATTTGCCCTCTCTTTGA 255
QY 1149 TCAATGCTTT 1158
DB 254 TCAATGCTTT 245

RESULT 11
A0875934/c
LOCUS V131C10 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomycetes
DEFINITION cerevisiae genomic 5', DNA sequence.
ACCESSION A0875934
VERSION A0875934.1 GI:6288178
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 796)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A., desEtages, S. A., Cheung, K. H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCTCTCTTTCTTTGGAGATAC
Class: transposon-tagged.
Location/Qualifiers 1..796

FEATURES
source
/organism="Saccharomyces cerevisiae"
/strain="Y2278 - S288C background, cir(0) rho(0)"
/db_xref="taxon:4932"
/clone_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"
/lab_host="E. coli"
/note="vector: pHS56-Sal; A yeast genomic DNA library without 2 micron or mitochondrial DNA was prepared in pHS56-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 218 a 167 c 209 g 202 t
ORIGIN

Query Match 10.9%; Score 126; DB 17; Length 796;

```

Best Local Similarity 100.0%; Pred. No. 5.6e-46;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 693 AACTCCATGGCAGCTCGCTTCGCTCCGCTCTTCATTAGAAATGTGGGAAAGGAT 752
Db 195 AACTCCATGGCAGCTCGCTTCGCTCCGCTCTTCATTAGAAATGTGGGAAAGGAT 136
QY 753 TTGGAGGAGAGACTCAACTCTATCGATCATGACATGAACAACAATAATTTGGTCTGGC 812
Db 135 TTGGAGGAGAGACTCAACTCTATCGATCATGACATGAACAACAATAATTTGGTCTGGC 76
QY 813 GAACATA 818
Db 75 GAACATA 70

RESULT 12
A0501433/c
LOCUS
DEFINITION V22C5 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
ACCESSION A0501433
VERSION A0501433.1 GI:4707083
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae
  ORGANISM Saccharomyces cerevisiae
    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
    Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
  REFERENCE 1 (bases 1 to 543)
  AUTHORS Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
    desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatis,D., Jansen,R.,
    Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
    Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
  TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
    Gene Disruption
  JOURNAL Unpublished (1999)
  COMMENT Contact: Kumar A
    Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
    Yale University
    P.O. Box 208103, New Haven, CT 06520-8103, USA
    Tel: 203 432 9949
    Fax: 203 432 6161
    Email: anuj.kumare@yale.edu
    te of mTn-3xHA/lacZ insertion.
    Seq primer: GGCCTCTCTTTTGGAGTAC
    Class: transposon-tagged.
  FEATURES
    source
      1..543
      /organism="Saccharomyces cerevisiae"
      /db_xref="taxon:4932"
      /clone_lib="mTn-3xHA/lacZ Insertion Library"
      /lab_host="E. coli"
      /note="vector: pHS6-Sal; A yeast genomic DNA library
      (lacking mitochondrial DNA) was prepared in pHS6-Sal;
      genomic DNA was size-fractionated (DNA of roughly 2-3 kb
      in length) prior to cloning. This library was
      subsequently mutagenized with a mTn-3xHA/lacZ
      minitransposon containing lacZ, URA3, and tet resistance."
    BASE COUNT 128 a 151 c 127 g 136 t
    ORIGIN

Query Match 6.6%; Score 76; DB 17; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.9e-23;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 GATCTTACGGTCTCACTAACTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATT 558
Db 140 GATCTTACGGTCTCACTAACTCTCTTCAACTGCTCAATAATTTCCCGCTATGCAAAATT 81
QY 559 CCCAAGACTACTTTTA 574
Db 80 CCCAAGACTACTTTTA 65

RESULT 14
A0502957/c
LOCUS
DEFINITION V47B7 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
ACCESSION A0502957
VERSION A0502957.1 GI:4705503
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae
  ORGANISM Saccharomyces cerevisiae
    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
    Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
  REFERENCE 1 (bases 1 to 525)
  AUTHORS Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 13:40:10 ; Search time 313 Seconds
(without alignments)
8331.674 Million cell updates/sec

Title: US-09-814-661a-1
Perfect score: 1158
Sequence: 1 aatgagcaacctgtcaaca.....cctctctgtatcaatgcttt 1158

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```
1: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1158	100.0	1158	21	Yeast SML1 (suppre
2	74	6.4	1767	17	Yeast calcineurin
3	45	3.9	65	21	PCR primer SML1-HI
4	45	3.9	65	21	PCR primer SML1-HI
5	35	3.0	4394	23	Drosophila melanog
6	32	2.8	11691	24	Human immune syste
7	31	2.7	53	22	Plasmodium falci
8	31	2.7	60	13	Microsatellite seq
9	31	2.7	128	13	Dinucleotide repea

10	31	2.7	128	13	AAQ32788	Microsatellite rep
11	31	2.7	128	15	AAQ57811	Repeat polymorphis
12	31	2.7	128	19	AAV15527	Polymorphic region
c 13	31	2.7	158	22	AAH93321	Plasmodium falci
c 14	31	2.7	191	22	AAH93328	Plasmodium falci
c 15	31	2.7	200	22	AAI61755	Soybean 515002 reg
16	31	2.7	297	18	AAAT91735	Mouse Smlm/CRP2 g
17	31	2.7	297	21	AAAC64711	Smlm/CRP2 related
18	31	2.7	297	22	AAAD10430	Intron 6 of mouse
c 19	31	2.7	445	21	ABN81094	Shrimp polynucleot
c 20	31	2.7	461	23	ABV54987	Human prostate exp
c 21	31	2.7	485	22	AAI88975	Human polynucleoti
c 22	31	2.7	1001	21	AAH51418	Human UGT1A7 relat
c 23	31	2.7	1014	8	AAH70770	Sequence encoding
c 24	31	2.7	1145	9	AAH80340	Sprague-Dawley rat
c 25	31	2.7	1264	15	AAQ67317	A. thaliana Ttl DN
c 26	31	2.7	1816	22	AAQ26168	Human type II ster
c 27	31	2.7	2437	16	AAQ97381	Human genital tiss
c 28	31	2.7	2437	18	AAAT99802	Human type II SRD5
c 29	31	2.7	2437	20	AAH88360	Human type II SRD5
c 30	31	2.7	2437	20	AAH88361	Human type II SRD5
c 31	31	2.7	2437	20	AAH88368	Human type II SRD5
c 32	31	2.7	2875	23	ABL08764	Drosophila melanog
c 33	31	2.7	3730	22	ABA19745	Human nervous syst
c 34	31	2.7	4195	21	AAA96418	Nucleotide sequenc
c 35	31	2.7	4388	18	AAAT60073	Selective marker g
c 36	31	2.7	4448	22	AAK76446	Human immune/haema
c 37	31	2.7	4625	23	ABL12644	Drosophila melanog
c 38	31	2.7	5089	23	ABL15200	Drosophila melanog
c 39	31	2.7	5400	21	AAA39732	Potato subclone pg
c 40	31	2.7	6020	24	ABL49307	Human polynucleoti
c 41	31	2.7	6020	24	ABL32198	Human immune syste
c 42	31	2.7	6039	23	ABL16342	Drosophila melanog
c 43	31	2.7	6227	24	ABA05235	Rice Reb b2ip enco
c 44	31	2.7	6298	24	ABL32434	Human immune syste
c 45	31	2.7	6568	24	ABL32447	Human immune syste

ALIGNMENTS

RESULT 1
AAAI4263
ID AAAI4263 standard; DNA; 1158 BP.
XX
AC AAAI4263;
XX
DT 15-AUG-2000 (first entry)
XX
DE Yeast SML1 (suppressor of mecl lethality) gene.
XX
KW SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT; ds.
XX
OS Saccharomyces cerevisiae.
XX
FH Key Location/Qualifiers
CDS 549..863
FT FT /*tag= a
FT FT /product= "Yeast Sml1 protein"
FT FT /function= "Inhibitor of ribonucleotide reductase"

WO200017225-A2.

30-MAR-2000.

24-SEP-1999; 99WO-US22260.

24-SEP-1998; 98US-0158858.

PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX Rothstein R, Zhao X;
 XX WPI: 2000-283539/24.
 DR P-PSDB; AAY90470.
 XX
 PT New Sm1 protein and its homologs, useful for treating cancer,
 PT microbial infection and ataxia telangiectasia and in screening for
 PT specific modulators
 XX
 PS Claim 6; Fig 1C; 98pp; English.
 XX
 CC This sequence represents the yeast SML1 (suppressor of mecl lethality)
 CC gene. This gene, located on chromosome XIII, encodes Sm1, an inhibitor
 CC of ribonucleotide reductase. The SML1 gene (assigned the name YML058w by
 CC the Stanford yeast genomic project) was initially isolated in a mecl-1
 CC strain by genetic analysis. Sm1 is a suppressor of the mecl mutant in
 CC yeast which is associated with abnormal levels of recombination in both
 CC meiosis and mitosis. Sm1 also permits cell growth in the absence of the
 CC cell cycle checkpoint proteins Mecl and Rad53, which, unlike most
 CC checkpoint proteins, are essential for cell growth. Sm1 binds to
 CC the largest subunit of ribonucleotide reductase (RNR) which catalyses the
 CC rate-limiting step of deoxyribonucleotide triphosphate (dNTP) synthesis,
 CC conversion of dNTPs to dNTPs. Binding of Sm1 reduces the activity of RNR
 CC and thus inhibits the synthesis of dNTPs. The SML1 gene, or fragments
 CC thereof may be used to isolate the human homologue of SML1. The Sm1
 CC protein, and compounds that modulate the interaction of Sm1 with
 CC ribonucleotide reductase (RNR), may be used to alter the rate at which
 CC cells divide. These are particularly useful for treating cancer,
 CC microbial infection and ataxia telangiectasia (AT), a condition in which
 CC patients are prone to cancer. Sm1 may also be used to screen for
 CC modulatory agents, to raise specific antibodies, and for stimulating the
 CC function of the ATM gene (a mammalian Mecl homologue which is mutated in
 CC AT). Anti-Sm1 antibodies are used as diagnostic and analytical
 CC immunoassay reagents and to remove Sm1 from serum or to titrate Sm1
 CC intracellularly.
 XX
 SQ Sequence 1158 BP; 289 A; 282 C; 196 G; 391 T; 0 other;
 Query Match 100.0%; Score 1158; DB 21; Length 1158;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AATGAGCAACCGTGTCAACAAGAGTGTCAAGACCGGCTACTATTCCCAAGGATCAGGT 60
 DB 1 AATGAGCAACCGTGTCAACAAGAGTGTCAAGACCGGCTACTATTCCCAAGGATCAGGT 60
 QY 61 TCCTTCTGCCAACATCATTCGCGTCAAGCTGCGGGCGGTCTTCTGACATTGGTAAGAA 120
 DB 61 TCCTTCTGCCAACATCATTCGCGTCAAGCTGCGGGCGGTCTTCTGACATTGGTAAGAA 120
 QY 121 TACTTCCCACTAAGAGCATGCTCTCTCTTTTGTAGGCAATGATAGGAAGAACAA 180
 DB 121 TACTTCCCACTAAGAGCATGCTCTCTCTTTTGTAGGCAATGATAGGAAGAACAA 180
 QY 181 TAGATTATAAATACGTCAAGATATAGTAGATATGTTTATGTTAGACCTCGTACATAG 240
 DB 181 TAGATTATAAATACGTCAAGATATAGTAGATATGTTTATGTTAGACCTCGTACATAG 240
 QY 241 GAATAAATGACGTTTTTTTTTGGGCAACATTTGAAATTTTTTTTGTACCTCGCGCTGA 300
 DB 241 GAATAAATGACGTTTTTTTTTGGGCAACATTTGAAATTTTTTTTGTACCTCGCGCTGA 300
 QY 301 GCCCAACGGGCTCCACTACCGCGGCTGCCATTTTGGGAAGTCAATCGTCCCAAAA 360
 DB 301 GCCCAACGGGCTCCACTACCGCGGCTGCCATTTTGGGAAGTCAATCGTCCCAAAA 360
 QY 361 AGGAAATAGCCATAAATATATGTTTGTACCTGTTTGGAAACATCGCCCGTTTCGCCGATTCG 420
 DB 361 AGGAAATAGCCATAAATATATGTTTGTACCTGTTTGGAAACATCGCCCGTTTCGCCGATTCG 420
 QY 421 CCTCAGCGGTATATAAAGAGATCTTTTTTTTCTCGCTGCTCCCTTCCATTTTAAATG 480

DB 421 CCTCAGCGGTATATAAAGAGATCTTTTTTTTCTCGCTGCTCCCTTCCATTTTAAATG 480
 QY 481 TCTTATCTGCTCCTTTTGTGATCTTACGGTCTCACTAAGCTCTCTTCAACTGCTCANTAT 540
 DB 481 TCTTATCTGCTCCTTTTGTGATCTTACGGTCTCACTAAGCTCTCTTCAACTGCTCANTAT 540
 QY 541 TTCCGGCTATGCAAAATTTCCCAAGACTACTTTTACGCTCAAAATGGCTGCCAACAAAC 600
 DB 541 TTCCGGCTATGCAAAATTTCCCAAGACTACTTTTACGCTCAAAATGGCTGCCAACAAAC 600
 QY 601 AAGCCCTTCCACATTCGCTACCGTACCGTACCGTACCGTACCGTACCGTACCGTACCG 660
 DB 601 AAGCCCTTCCACATTCGCTACCGTACCGTACCGTACCGTACCGTACCGTACCGTACCG 660
 QY 661 CTATGGCTGAGGTTCTTATGTTCTTACTCAAAACTCCATGGGAGCTCCGCTTCTGGCT 720
 DB 661 CTATGGCTGAGGTTCTTATGTTCTTACTCAAAACTCCATGGGAGCTCCGCTTCTGGCT 720
 QY 721 CCGCTTCTTCAATAGAAATGTGGAAAGAGTGTGGAGGAGAGACTCAACTCTATCGATC 780
 DB 721 CCGCTTCTTCAATAGAAATGTGGAAAGAGTGTGGAGGAGAGACTCAACTCTATCGATC 780
 QY 781 ATGACATGAACAACAATAATTTGGTCTTGGCGAACTAAAATCTATGTTCAACAGGCTA 840
 DB 781 ATGACATGAACAACAATAATTTGGTCTTGGCGAACTAAAATCTATGTTCAACAGGCTA 840
 QY 841 AGGTCGAGGAATGAGACTTCTTAAAGTTCCTTCACTACTCTTTCTTCTTCTTCTTCT 900
 DB 841 AGGTCGAGGAATGAGACTTCTTAAAGTTCCTTCACTACTCTTTCTTCTTCTTCTTCT 900
 QY 901 CCCACTAGTCTGTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960
 DB 901 CCCACTAGTCTGTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960
 QY 961 CTATTTGTTGTCATCTCGAAACATCTCTCCCGTGCAATTTTCTTCTTCTTCTTCTTCT 1020
 DB 961 CTATTTGTTGTCATCTCGAAACATCTCTCCCGTGCAATTTTCTTCTTCTTCTTCTTCT 1020
 QY 1021 AT 1080
 DB 1021 AT 1080
 QY 1081 AAAGATAGTCTATATAGCTTTGATACAGCTAGATATCGCTAGCGCCCAACATTTGTCGCC 1140
 DB 1081 AAAGATAGTCTATATAGCTTTGATACAGCTAGATATCGCTAGCGCCCAACATTTGTCGCC 1140
 QY 1141 TCTCTTGAATCAATGCTTT 1158
 DB 1141 TCTCTTGAATCAATGCTTT 1158
 RESULT 2
 AAT27106
 ID AAT27106 standard; DNA; 1767 BP.
 XX
 AC AAT27106;
 XX
 DT 27-OCT-1996 (first entry)
 XX
 DE Yeast calcineurin subunit CNA2deltaC DNA.
 XX
 KW Calcineurin; CNA2deltaC; Immunosuppressive; Immunosuppressant;
 KW gene therapy; ds.
 XX
 OS Yeast.
 XX
 FH Key Location/Qualifiers
 FT CDS 262..1767
 FT /*tag= a
 XX
 PN W09612806-A1.
 XX

```

PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US13580.
XX
XX 24-OCT-1994; 94US-0328322.
XX
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI Cyert MS, Huang L;
XX
DR WPI: 1996-230607/23.
XX
DR P-PSDB: AAR96087.
XX
PT Calcineurin interacting polypeptide - enhances immunosuppressive
PT effects of calcineurin-targeted immunosuppressant by potentiating
PT interaction between an immunophilin and CN
XX
PS Disclosure; Page 69-71; 90pp; English.
XX
CC This DNA encodes a subunit of calcineurin, designated CNA2deltaC,
CC and may be used to form a fusion protein which consists of this CN
CC subunit and a calcineurin-interactin polypeptide. The fusion
CC protein may be used in a method for the identification of a small
CC molecule immunosuppressant compound.
XX
SQ Sequence 1767 BP; 524 A; 416 C; 353 G; 474 T; 0 other;
Query Match 6.4%; Score 74; DB 17; Length 1767;
Best Local Similarity 100.0%; Pred. No. 3.8e-25;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1085 ATAGCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTC 1144
DB 1 ATAGCTATATACGTTTGATACAGCTAGATATCGTAGCGCCCAACATTGTCCCCCTC 60
QY 1145 TTGATCAATGCTTT 1158
DB 61 TTGATCAATGCTTT 74
RESULT 3
AAAL4268
ID AAAL4268 standard; DNA; 65 BP.
XX
AC AAAL4268;
XX
DT 15-AUG-2000 (first entry)
XX
DE PCR primer SML1-HIS3 5', used in the study of the yeast SML1 gene.
XX
KW SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT; PCR primer; ss.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200017225-A2.
XX
PD 30-MAR-2000.
XX
PF 24-SEP-1999; 99WO-US22260.
XX
PR 24-SEP-1998; 98US-0158858.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Rothstein R, Zhao X;
XX
DR WPI: 2000-283539/24.
XX
PT New Sml1 protein and its homologs, useful for treating cancer,

```

```

PT microbial infection and ataxia telangiectasia and in screening for
PT specific modulators
XX
PS Disclosure; Page 65; 98pp; English.
XX
CC The invention relates to the yeast SML1 (suppressor of mecl lethality)
CC gene (AAAL4263) and to the Sml1 protein (AA90470). The Sml1 gene,
CC located on chromosome XIII, encodes Sml1, an inhibitor of ribonucleotide
CC reductase. The SML1 gene (assigned the name YML058w by the Stanford yeast
CC genomic project) was initially isolated in a mecl-1 strain by genetic
CC analysis. Sml1 is a suppressor of the mecl mutant in yeast which is
CC associated with abnormal levels of recombination in both meiosis and
CC mitosis. Sml1 also permits cell growth in the absence of the cell cycle
CC checkpoint proteins Mec1 and Rad53, which, unlike most checkpoint
CC proteins, are essential for cell growth. Sml1 binds to the largest
CC subunit of ribonucleotide reductase (RNR) which catalyses the
CC rate-limiting step of deoxynucleotide triphosphate (dNTP) synthesis.
CC conversion of NTPs to dNTPs. Binding of Sml1 reduces the activity of RNR
CC and thus inhibits the synthesis of dNTPs. The SML1 gene, or fragments
CC thereof may be used to isolate the human homologue of SML1. The Sml1
CC protein, and compounds that modulate the interaction of Sml1 with
CC ribonucleotide reductase (RNR), may be used to alter the rate at which
CC cells divide. These are particularly useful for treating cancer, in
CC which microbial infection and ataxia telangiectasia (AT), a condition in which
CC patients are prone to cancer. Sml1 may also be used to screen for
CC modulatory agents, to raise specific antibodies, and for stimulating the
CC function of the ATM gene (a mammalian Mec1 homologue which is mutated in
CC AT). Anti-Sml1 antibodies are used as diagnostic and analytical
CC immunoassay reagents and to remove Sml1 from serum or to titrate Sml1
CC intracellularly. Sequences AAA24266-AL4285 represent PCR primers used in
CC the study of the yeast SML1 gene.
XX
SQ Sequence 65 BP; 12 A; 23 C; 11 G; 19 T; 0 other;
Query Match 3.9%; Score 45; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 2.9e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 502 CTTACGGTCTCACTAACCTCTCTCACTGCTCAATAATTTCCCG 546
DB 1 CTTACGGTCTCACTAACCTCTCTCACTGCTCAATAATTTCCCG 45
RESULT 4
AAAL4269/C
ID AAAL4269 standard; DNA; 65 BP.
XX
AC AAAL4269;
XX
DT 15-AUG-2000 (first entry)
XX
DE PCR primer SML1-HIS3 3', used in the study of the yeast SML1 gene.
XX
KW SML1 gene; YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT; PCR primer; ss.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200017225-A2.
XX
PD 30-MAR-2000.
XX
PF 24-SEP-1999; 99WO-US22260.
XX
PR 24-SEP-1998; 98US-0158858.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Rothstein R, Zhao X;
XX

```


PS Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX

SQ Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;

Query Match 2.88; Score 32; DB 24; Length 11691;

Best Local Similarity 100.0%; Pred. No. 3.7e-05;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1011 TTATACATATATATATATATATATATATATATAT 1042

Db 442 TTATACATATATATATATATATATATATATATAT 411

RESULT 7

AAH93337/C

ID AAH93337 standard; DNA; 53 BP.

XX

AC AAH93337;

XX

DT 04-OCT-2001 (first entry)

XX

DE Plasmodium falciparum MAL3PC polynucleotide SEQ ID NO 58.

XX

XX Human; antisense-therapy; gene-therapy; diagnostic; forensic;

KW gene mapping; ds.

KW

XX Plasmodium falciparum.

OS

XX WO200152616-A2.

PN

XX 26-JUL-2001.

PD

XX

PF 22-DEC-2000; 2000WO-US35190.

XX

PR 23-DEC-1999; 99US-0471275.

PR

PR 21-JAN-2000; 2000US-0488725.

PR

XX 25-APR-2000; 2000US-0552317.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-451890/48.

XX

PT Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -

XX

PS Example 4; Page 121; 135pp; English.

XX

CC The invention relates to an isolated human polynucleotide (AAH75398)

CC encoding a novel polypeptide (AAG64527) useful in antisense-therapy and

CC gene-therapy, in diagnostics, forensics, gene mapping and identification

CC of mutations responsible for genetic disorders and other traits.

CC Polynucleotide sequences with potential homology were also identified

CC (AAH93283-AAH93356).

XX

SQ Sequence 53 BP; 29 A; 0 C; 1 G; 23 T; 0 other;

Query Match 2.78; Score 31; DB 22; Length 53;

Best Local Similarity 100.0%; Pred. No. 0.00014;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1012 TATATACATATATATATATATATATATATATAT 1042

Db TATATACATATATATATATATATATATATATATATAT 59

RESULT 9

AAQ22193

ID AAQ22193 standard; DNA; 128 BP.

XX

AC AAQ22193;

XX

DT 17-DEC-2001 (updated)

DT 26-JUN-1992 (first entry)

Db 53 TATATACATATATATATATATATATATATATATAT 23

RESULT 8

AAQ33792

ID AAQ33792 standard; DNA; 60 BP.

XX

AC AAQ33792;

XX

DT 02-FEB-1993 (first entry)

XX

DE Microsatellite sequence from clone TGLA20.

XX

KW PCR; selection; primers; OPTIPRIM; breeding; cattle; parentage;

KW genetic mapping; traits; amplification; ss.

XX

OS Bos taurus.

XX

PN WO9213102-A.

XX

PD 06-AUG-1992.

XX

PF 15-JAN-1992; 92WO-US00340.

XX

PR 15-JAN-1991; 91US-0642342.

XX

PA (GENM-) GENMARK.

XX

PI Georges M, Massey JM;

XX

DR WPI; 1992-284684/34.

XX

PT Polymorphic bovine DNA markers - used in genetic identification,

PT gene mapping, and selective breeding

XX

PS Table 7; Page 246; 517pp; English.

XX

CC The sequence is that of a bovine microsatellite sequence obtd. by

CC screening a library of bovine MboI DNA fragments of between

CC 250 and 500 bp with an (AC)₁₅ and a (TC)₁₅ oligonucleotide probe.

CC One out of 50 clones cross-hybridised. Assuming independent

CC distribution of microsatellites and MboI sites, the frequency of

CC (T5)n > 9 microsatellites in the bovine genome is estimated at >100,

CC 000. The sequence information for ca. 230 such bovine microsatellites

CC is summarised in the specification and indexed herein (see below).

CC The sequences upstream and downstream of the microsatellite sequence

CC were used to generate the required PCR primers for in vitro

CC amplification of the corresp. microsatellite (using the program

CC OPTIPRIM). The microsatellites may be used to identify individuals,

CC for parentage testing, and in the genetic mapping of economic trait

CC loci, or genes involved in the determination of economically important

CC traits esp. in cattle, to allow selective breeding.

CC See also AAQ33501-34437.

XX

SQ Sequence 60 BP; 18 A; 1 C; 12 G; 29 T; 0 other;

Query Match 2.78; Score 31; DB 13; Length 60;

Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1012 TATATACATATATATATATATATATATATATAT 1042

Db TATATACATATATATATATATATATATATATATATAT 59

RESULT 9

AAQ22193

ID AAQ22193 standard; DNA; 128 BP.

XX

AC AAQ22193;

XX

DT 17-DEC-2001 (updated)

DT 26-JUN-1992 (first entry)

XX DE Dinucleotide repeat polymorphic DNA marker.
 XX KW Primer: PCR; amplification: individualisation; forensic; paternity;
 XX KW prenatal; screening; genetic mapping; ss.
 XX OS Homo sapiens.
 XX PN USN7707501-N.
 XX PD 28-JAN-1992.
 XX PF 29-MAY-1991; 91US-0707501.
 XX PR 29-MAY-1991; 91US-0707501.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX PI Polymeropoulos MH, Merrill C;
 XX DR WPI; 1992-088617/11.
 XX PT New polymorphic DNA markers - used for human sample
 XX PT individualisation in forensic, paternity and prenatal screening
 XX PS and in genetic mapping
 XX PS Disclosure; Fig 8; 22pp; English.
 XX CC The dinucleotide repeat polymorphism marker is accompanied by
 XX CC beginning and ending nucleotide sequences. Since a polymorphic
 XX CC marker and an index locus occur as a pair, attaching a primer oligo-
 XX CC nucleotide to the polymorphic marker allows PCR amplification of the
 XX CC segment pair. The markers have high polymorphism information content
 XX CC (PIC) values and can be used for human sample individualisation.
 XX CC Conducting three sets of PCR on the same DNA segment samples while
 XX CC using a different PCR primer pair yields accurate and informative
 XX CC test results. The assay is simple to perform and requires only small
 XX CC amounts of genomic DNA, with results being obtained within 24 hours.
 XX CC For example, 60 ng of genomic DNA was used as a template for PCR
 XX CC with 80 ng of each oligonucleotide primer under standard conditions.
 XX CC Each PCR mixture was electrophoresed by PAGE sequencing gels and
 XX CC visualised by autoradiography. The assays can detect very small
 XX CC differences in nucleotide sequences and can be used in forensic
 XX CC screening, in paternity and prenatal screening and in genetic mapping.
 XX CC See also AAQ22186-94.
 XX CC (Note: Revised entry submitted to correct the patent number format of
 XX CC US Government-owned NTIS applications to prevent clashes with ongoing US
 XX CC granted patent numbers. For further information please visit the Derwent
 XX CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
 XX SQ Sequence 128 BP; 48 A; 9 C; 16 G; 55 T; 0 other;
 Query Match 2.7%; Score 31; DB 13; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1012 TATATACATATATATATATATATATATAT 1042
 DB 28 TATATACATATATATATATATATATATATAT 58
 RESULT 10
 AAQ32788
 ID AAQ32788 standard; DNA; 128 BP.
 XX AC AAQ32788;
 XX DT 05-MAY-1993 (first entry)
 XX DE Microsatellite repeat polymorphism DNA marker.
 XX KW PIC; high polymorphism information content; forensic; screening;
 XX KW genetic mapping; paternity; prenatal.
 Query Match 2.7%; Score 31; DB 13; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1012 TATATACATATATATATATATATATATAT 1042
 DB 28 TATATACATATATATATATATATATATATAT 58
 RESULT 11
 AAQ57811
 ID AAQ57811 standard; DNA; 128 BP.
 XX AC AAQ57811;
 XX DT 21-AUG-1994 (first entry)
 XX DE Repeat polymorphism marker sequence #2.
 XX KW Primer: assay; subtle difference; dinucleotide; tetranucleotide;
 XX KW repeat; polymorphism; PCR; polymerase chain reaction; amplify; PAGE;
 XX KW autoradiography; migration pattern; length variation; genetic mapping;
 XX KW forensic screening; paternity; prenatal; screening; microsatellite;
 XX KW human; ss.
 XX OS Synthetic.
 XX PN WO9403640-A.
 XX PD 17-FEB-1994.
 XX PF 30-JUL-1993; 93WO-US07183.
 XX PR 31-JUL-1992; 92US-0922723.
 XX PR 28-SEP-1992; 92US-0952277.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Merrill CR, Polymeropoulos MH;
 XX DR WPI; 1994-065727/08.
 XX KW

XX OS Synthetic.
 XX PN WO9221693-A.
 XX PD 10-DEC-1992.
 XX PF 27-MAY-1992; 92WO-US04195.
 XX PR 29-MAY-1991; 91US-0707501.
 XX PR 27-NOV-1991; 91US-0799828.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX PI Merrill CR, Polymeropoulos MH;
 XX DR WPI; 1992-433606/52.
 XX CC Oligo-nucleotide primers for polymerase chain reaction
 XX CC amplification - which detect DNA polymorphisms and are useful for
 XX CC prenatal and paternity screening, and genetic mapping
 XX PS Disclosure; Fig 8; 44pp; English.
 XX CC The sequence is that of a microsatellite repeat polymorphism marker
 XX CC sequence which can be used for human individualisation, in forensic
 XX CC screening, in paternity and prenatal screening as well as in genetic
 XX CC mapping.
 XX SQ Sequence 128 BP; 48 A; 9 C; 16 G; 55 T; 0 other;
 Query Match 2.7%; Score 31; DB 13; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1012 TATATACATATATATATATATATATATAT 1042
 DB 28 TATATACATATATATATATATATATATATAT 58
 RESULT 11
 AAQ57811
 ID AAQ57811 standard; DNA; 128 BP.
 XX AC AAQ57811;
 XX DT 21-AUG-1994 (first entry)
 XX DE Repeat polymorphism marker sequence #2.
 XX KW Primer: assay; subtle difference; dinucleotide; tetranucleotide;
 XX KW repeat; polymorphism; PCR; polymerase chain reaction; amplify; PAGE;
 XX KW autoradiography; migration pattern; length variation; genetic mapping;
 XX KW forensic screening; paternity; prenatal; screening; microsatellite;
 XX KW human; ss.
 XX OS Synthetic.
 XX PN WO9403640-A.
 XX PD 17-FEB-1994.
 XX PF 30-JUL-1993; 93WO-US07183.
 XX PR 31-JUL-1992; 92US-0922723.
 XX PR 28-SEP-1992; 92US-0952277.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Merrill CR, Polymeropoulos MH;
 XX DR WPI; 1994-065727/08.
 XX KW

PT New polynucleotide sequences - derived from polymorphic
PT micro:satellite repeats, used for characterising human
PT individuals for forensic, paternity and prenatal screening and
PT genetic mapping

PS Disclosure: Page 33; 72pp; English.

XX The sequences given in AA057782-866 are primers which were used in
CC an assay for measuring the subtle differences in genetic material
CC regarding an added or omitted set of dinucleotide or tetranucleotide
CC repeat polymorphisms. The method comprises obtaining polynucleotide
CC segments comprising the repeat polymorphisms in an amount effective
CC for testing and amplifying the segments by a PCR procedure using a
CC pair of oligonucleotide primers capable of amplifying the polymorphism
CC containing sequence. The amplified sequences are resolved using PAGE
CC and the resolved sequences are compared by autoradiography to observe
CC the differences in migration pattern due to length variation. The
CC polynucleotides provide a fast and accurate test for measuring the
CC subtle differences in individuals in eg. forensic screening, paternity
CC and prenatal screening and genetic mapping. The polynucleotides are
CC specific for polymorphic microsatellite repeats based on previously
CC sequenced human genes.

XX Sequence 128 BP; 48 A; 9 C; 16 G; 55 T; 0 other;

Query Match 2.7%; Score 31; DB 15; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATATATATAT 1042
DB 28 TATATACATATATATATATATATATATATATATAT 58

RESULT 12

AAV15527
ID AAV15527 standard; DNA; 128 BP.

AC AAV15527;

DT 22-MAY-1998 (first entry)

DE Polymorphic region of human genome.

XX PCR primer; polymorphic region; human genome; dinucleotide repeat;
KW tetranucleotide repeat; assaying; forensic analysis;
KW paternity determination; prenatal screening; genetic mapping; ss.

OS Homo sapiens.

PN US57211100-A.

PD 24-FEB-1998.

XX 07-JUN-1995; 95US-0480366.

XX 29-MAY-1991; 91US-0707501.

PR 09-JUN-1993; 93US-0074275.

PR 07-JUN-1995; 95US-0480366.

XX (USHS) US DEPT HEALTH & HUMAN SERVICES.

XX Merrill CR, Polymeropoulos MH;

DR WPI; 1998-168402/15.

PT Oligo:nucleotide primers - are useful for human gene polymorphism
PT analysis

PS Claim 9; Columns 11-12; 10pp; English.

XX The oligonucleotides AAV15520-25 are useful as primers for the PCR
CC amplification of polymorphic regions of the human genome having

CC characteristic dinucleotide or tetranucleotide repeats, e.g.
CC AAV15526-28. They are especially useful in an assay for correlating
CC the source of a test nucleic acid sample with an individual,
CC comprising amplifying a nucleic acid sample from the individual
CC using the primer pair AAV15520/21, AAV15522/23 or AAV15524/25,
CC amplifying the test sample using the same primer pair, resolving the
CC amplified products by gel electrophoresis and comparing the mobilities
CC of the amplification products from the two samples. The assay can be
CC used for forensic analysis, determining paternity, prenatal
CC screening and genetic mapping. The primer pairs have high
CC polymorphism information content (PIC) values of 0.91-0.92.

XX Sequence 128 BP; 48 A; 9 C; 16 G; 55 T; 0 other;

Query Match 2.7%; Score 31; DB 19; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATATATATAT 1042
DB 28 TATATACATATATATATATATATATATATATATATAT 58

RESULT 13

AAH93321/c

ID AAH93321 standard; DNA; 158 BP.

AC AAH93321;

DT 04-OCT-2001 (first entry)

DE Plasmodium falciparum MAL3PC polynucleotide SEQ ID NO 42.

KW Human; antisense-therapy; gene-therapy; diagnostic; forensic;
KW gene mapping; ds.

OS Plasmodium falciparum.

PN WO200152616-A2.

PD 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35190.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451890/48.

PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -

PS Example 4; Page 118; 135pp; English.

XX The invention relates to an isolated human polynucleotide (AAH75398)
CC encoding a novel polypeptide (AAG64527) useful in antisense-therapy and
CC gene-therapy, in diagnostics, forensics, gene mapping and identification
CC of mutations responsible for genetic disorders and other traits.
CC Polynucleotide sequences with potential homology were also identified
CC (AAH93283-AAH93356).

XX Sequence 158 BP; 77 A; 12 C; 9 G; 60 T; 0 other;

Query Match 2.7%; Score 31; DB 22; Length 158;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATATATATAT 1042

```

Db      50 TATATACATATATATATATATATATATATAT 20
|||||
RESULT 14
AAH93328/c
ID   AAH93328 standard; DNA; 191 BP.
XX
XX
AC   AAH93328;
XX
XX
DT   04-OCT-2001 (first entry)
XX
XX
DE   Plasmodium falciparum MAL3PC polynucleotide SEQ ID NO 49.
XX
XX
KW   Human; antisense-therapy; gene-therapy; diagnostic; forensic;
KW   gene mapping; ds.
XX
XX
OS   Plasmodium falciparum.
XX
XX
PN   WO200152616-A2.
XX
XX
PD   26-JUL-2001.
XX
XX
PF   22-DEC-2000; 2000WO-US35190.
XX
XX
PR   23-DEC-1999; 99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
XX
PA   (HYSE-) HYSEQ INC.
XX
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
XX
DR   WPI; 2001-451890/48.
XX
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
XX
PS   Example 4; Page 119-120; 135pp; English.
XX
XX
CC   The invention relates to an isolated human polynucleotide (AAH75398)
CC   encoding a novel polypeptide (AA664527) useful in antisense-therapy and
CC   gene-therapy, in diagnostics, forensics, gene mapping and identification
CC   of mutations responsible for genetic disorders and other traits.
CC   Polynucleotide sequences with potential homology were also identified
CC   (AAH93283-AAH93356).
XX
XX
SQ   Sequence 191 BP; 89 A; 5 C; 12 G; 85 T; 0 other;

Query Match      2.7%; Score 31; DB 22; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY   1012 TATATACATATATATATATATATATATATAT 1042
      |||||||
Db    169 TATATACATATATATATATATATATATATAT 139

RESULT 15
AAI61755/G
ID   AAI61755 standard; DNA; 200 BP.
XX
XX
AC   AAI61755;
XX
XX
DT   16-OCT-2001 (first entry)
XX
XX
DE   Soybean 515002 region G2 DNA, SEQ ID NO: 386.
XX
XX
KW   Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
XX
OS   Glycine max.

```

```

XX
PN   WO200151627-A2.
XX
XX
PD   19-JUL-2001.
XX
XX
PF   05-JAN-2001; 2001WO-US00552.
XX
XX
PR   07-JAN-2000; 2000US-0174880.
XX
XX
PA   (MONS ) MONSANTO CO.
XX
XX
PI   Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
XX
DR   WPI; 2001-425872/45.
XX
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs -
XX
XX
PS   Claim 25; Page 1132; 1353pp; English.
XX
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule
CC   provided in the specification.
XX
XX
SQ   Sequence 200 BP; 66 A; 17 C; 31 G; 86 T; 0 other;

Query Match      2.7%; Score 31; DB 22; Length 200;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY   1012 TATATACATATATATATATATATATATATAT 1042
      |||||||
Db    172 TATATACATATATATATATATATATATATAT 142

Search completed: July 17, 2003, 15:45:14
Job time : 316 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 15:34:49 ; Search time 84 Seconds
(without alignments)
4227.758 Million cell updates/sec

Title: US-09-814-661A-1
Perfect score: 1158
Sequence: 1 aatgagcaaccgtgtcaaca.....cctctcttgatgaatgcttt 1158

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	6.4	1767	1	US-08-328-322-20
2	74	6.4	2353	1	US-08-328-322-11
3	31	2.7	128	1	US-07-922-723A-8
4	31	2.7	128	1	US-07-799-828C-8
5	31	2.7	128	1	US-08-074-275-8
6	31	2.7	128	1	US-08-480-366-8
7	31	2.7	128	2	US-07-952-277A-8
8	31	2.7	297	1	US-08-616-368A-17
9	31	2.7	297	3	US-09-054-298-17
10	31	2.7	297	4	US-08-818-655-17
11	31	2.7	1160	1	US-08-006-082A-4
12	31	2.7	1264	1	US-08-006-082A-1
13	31	2.7	2437	1	US-07-795-859B-5
14	31	2.7	2437	1	US-08-457-816-5
15	31	2.7	2437	4	US-09-235-538-1
16	30	2.6	43795	3	US-08-742-185-101
17	29	2.5	6138	4	US-09-067-800-4
18	29	2.5	6138	4	US-09-349-677-4
19	29	2.5	162450	4	US-09-345-882-1
20	29	2.5	246240	2	US-08-724-394A-20
21	29	2.5	246240	2	US-08-724-394A-21
22	29	2.5	246240	2	US-08-724-394A-22
23	28	2.4	1265	1	US-08-182-060A-5
24	28	2.4	1265	1	US-08-712-702A-5
25	28	2.4	2403	1	US-08-441-139-17
26	28	2.4	19124	2	US-08-487-826B-13
27	27	2.3	804	1	US-08-126-593A-3

28	27	2.3	804	1	US-08-454-039A-3	Sequence 3, Appli
29	27	2.3	1906	1	US-08-207-904-18	Sequence 18, Appl
30	27	2.3	2551	1	US-08-486-013-70	Sequence 70, Appl
31	27	2.3	2551	2	US-08-482-279-70	Sequence 70, Appl
32	27	2.3	2551	2	US-08-342-288-70	Sequence 70, Appl
33	27	2.3	2551	3	US-09-015-968-70	Sequence 70, Appl
34	27	2.3	2551	4	US-09-397-386-70	Sequence 70, Appl
35	27	2.3	3072	4	US-09-522-217-55	Sequence 55, Appl
36	27	2.3	3072	4	US-09-522-217-55	Sequence 55, Appl
37	27	2.3	6152	4	US-08-973-462-1	Sequence 1, Appl
38	26	2.2	128	1	US-07-922-723A-8	Sequence 8, Appl
39	26	2.2	128	1	US-07-799-828C-8	Sequence 8, Appl
40	26	2.2	128	1	US-08-074-275-8	Sequence 8, Appl
41	26	2.2	128	1	US-08-480-366-8	Sequence 8, Appl
42	26	2.2	128	2	US-07-952-277A-8	Sequence 8, Appl
43	26	2.2	297	1	US-08-616-368A-17	Sequence 17, Appl
44	26	2.2	297	3	US-09-054-298-17	Sequence 17, Appl
45	26	2.2	297	4	US-08-818-655-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-328-322-20
; Sequence 20, Application US/08328322
; Patent No. 5723436
; GENERAL INFORMATION:
; APPLICANT: Huang, Laiqiang
; TITLE OF INVENTION: Calcineurin Interacting Protein Compositions
; TITLE OF INVENTION: and Methods
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,322
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: P38,615
; REFERENCE/DOCKET NUMBER: 8600-0151.10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0960
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1767 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: DNA fragment containing
; INDIVIDUAL ISOLATE: CNA2deltaC coding sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262..1767
US-08-328-322-20

QY 1145 TTGATCAATGCTTT 1158
|||||

```

; ZIP: 22314
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/799,828C
; FILING DATE: 19911127
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: D.J. Mills
; REGISTRATION NUMBER: 34,506
; REFERENCE/DOCKET NUMBER: 717081A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-799-828C-8

Query Match 2.7%; Score 31; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATAT 1042
|||||
Db 28 TATATACATATATATATATATATATATAT 58

RESULT 5
US-08-074-275-8
; Sequence 8, Application US/08074275
; Patent No. 5468610
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,275
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/707,501
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

US-08-074-275-8

Query Match 2.7%; Score 31; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATAT 1042
|||||
Db 28 TATATACATATATATATATATATATATAT 58

RESULT 6
US-08-480-366-8
; Sequence 8, Application US/08480366
; Patent No. 5721100
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,366
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-480-366-8

Query Match 2.7%; Score 31; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATAT 1042
|||||
Db 28 TATATACATATATATATATATATATATAT 58

RESULT 7
US-07-952-277A-8
; Sequence 8, Application US/07952277A
; Patent No. 5861504
; GENERAL INFORMATION:
; APPLICANT: Drs. Michael H. Polymeropoulos
; APPLICANT: and Carl R. Merrill
; TITLE OF INVENTION: ELEVEN HIGHLY INFORMATIVE
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker

STREET: Suite 300, 99 Canal Center Plaza
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: DOS Text File
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/952,277A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: D.J. Mills
REGISTRATION NUMBER: 34506
REFERENCE/DOCKET NUMBER: 717081C
TELEPHONE: 703 684 1111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 128
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-952-277A-8

Query Match 2.7%; Score 31; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATAT 1042
DB 28 TATATACATATATATATATATATATAT 58

RESULT 8
US-08-616-368A-17
Sequence 17, Application US/08616368A
Patent No. 5767262
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Haber, Edgar
APPLICANT: Jain, Mukesh
APPLICANT: Yet, Shaw-Fang
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/616,368A
FILING DATE: 15-MAR-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/022001
TELEPHONE: 617/542-8906
TELEFAX: 617/542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-616-368A-17

Query Match 2.7%; Score 31; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATAT 1042
DB 232 TATATACATATATATATATATATATATAT 262

RESULT 9
US-09-054-298-17
Sequence 17, Application US/05054298
Patent No. 6136953
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Haber, Edgar
APPLICANT: Jain, Mukesh
APPLICANT: Yet, Shaw-Fang
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,298
FILING DATE: 02-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/616,368
FILING DATE: 15-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Beattie, Ph.D., Ingrid A.
REGISTRATION NUMBER: P-42,306
REFERENCE/DOCKET NUMBER: 05433/022002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-054-298-17

Query Match 2.7%; Score 31; DB 3; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATAT 1042
DB 232 TATATACATATATATATATATATATATAT 262

RESULT 10
US-08-818-655-17
; Sequence 17, Application US/08818655
; Patent No. 6258557
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-Pn
; APPLICANT: Haber, Edgar
; APPLICANT: Jain, Mukesh
; APPLICANT: Yet, Shaw-Fang
; TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,655
; FILING DATE: 14-MAR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/616,368
; FILING DATE: 15-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/030001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-818-655-17

Query Match 2.7%; Score 31; DB 4; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1012 TATATACATATATATATATATATAT 1042
Db 232 TATATACATATATATATATATATAT 262

RESULT 11
US-08-006-082A-4/c
; Sequence 4, Application US/08006082A
; Patent No. 5489743
; GENERAL INFORMATION:
; APPLICANT: ROBINSON ET AL.
; TITLE OF INVENTION: TRANSGENIC MAMMALS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: Amgen Center
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in., DS, 2.0 MB

; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 7.0.
; SOFTWARE: Microsoft Word Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/006,082A
; FILING DATE: 19-JAN-1993
; CLASSIFICATION: 800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1160 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single stranded
; TOPOLOGY: Linear
US-08-006-082A-4

Query Match 2.7%; Score 31; DB 1; Length 1160;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1012 TATATACATATATATATATATATAT 1042
Db 121 TATATACATATATATATATATATATAT 91

RESULT 12
US-08-006-082A-1/c
; Sequence 1, Application US/08006082A
; Patent No. 5489743
; GENERAL INFORMATION:
; APPLICANT: ROBINSON ET AL.
; TITLE OF INVENTION: TRANSGENIC MAMMALS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: Amgen Center
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in., DS, 2.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 7.0.
; SOFTWARE: Microsoft Word Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/006,082A
; FILING DATE: 19-JAN-1993
; CLASSIFICATION: 800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1264 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single stranded
; TOPOLOGY: Linear
US-08-006-082A-1

Query Match 2.7%; Score 31; DB 1; Length 1264;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1012 TATATACATATATATATATATATAT 1042
Db 187 TATATACATATATATATATATATATAT 157

RESULT 13
US-07-795-859B-5/c
; Sequence 5, Application US/07795859B
; Patent No. 5422262
; GENERAL INFORMATION:
; APPLICANT: Anderson, Stefan
; APPLICANT: Russell, David W.

;; TITLE OF INVENTION: Steroid 5'-Reductases
;; NUMBER OF SEQUENCES: 37
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Arnold, White and Durkee
;; STREET: P.O. Box 4433
;; CITY: Houston
;; STATE: TX
;; COUNTRY: USA
;; ZIP: 77210
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: ASCII-DOS
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/795.859B
;; FILING DATE: 18-NOV-1991
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Parker, David L.
;; REGISTRATION NUMBER: 32,165
;; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (512) 320-7200
;; TELEFAX: (512) 474-7677
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2437 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 28..789
;;
US-07-795-859B-5

Query Match 2.7%; Score 31; DB 1; Length 2437;
Best Local Similarity 100.0%; Pred. No. 1.le-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1012 TATATACATATATATATATATATATATAT 1042
|||||
Db 926 TATATACATATATATATATATATATATAT 896

RESULT 14
US-08-457-616-5/c
;; Sequence 5, Application US/08457616
;; Patent No. 5679521
;; GENERAL INFORMATION:
;; APPLICANT: Anderson, Stefan
;; APPLICANT: Russell, David W.
;; TITLE OF INVENTION: Steroid 5'-Reductases
;; NUMBER OF SEQUENCES: 37
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Arnold, White and Durkee
;; STREET: P.O. Box 4433
;; CITY: Houston
;; STATE: TX
;; COUNTRY: USA
;; ZIP: 77210
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: ASCII-DOS
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/457.616
;; FILING DATE: 01-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/795.859

;; FILING DATE: 18-NOV-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Parker, David L.
;; REGISTRATION NUMBER: 32,165
;; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (512) 320-7200
;; TELEFAX: (512) 474-7677
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2437 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 28..789
;;
US-08-457-616-5

Query Match 2.7%; Score 31; DB 1; Length 2437;
Best Local Similarity 100.0%; Pred. No. 1.le-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1012 TATATACATATATATATATATATATATAT 1042
|||||
Db 926 TATATACATATATATATATATATATATAT 896

RESULT 15
US-09-235-538-1/c
;; Sequence 1, Application US/09235538
;; Patent No. 6395479
;; GENERAL INFORMATION:
;; APPLICANT: Reichardt, Juergen, K.V., Ph.D.
;; APPLICANT: Gerhardt, Coetzee, A., Ph.D.
;; APPLICANT: Henderson, Brian E., M.D.
;; APPLICANT: Makridakis, Nick
;; APPLICANT: Ross, Ronald M.D.
;; APPLICANT: University of Southern California
;; TITLE OF INVENTION: ANDROGEN-METABOLIC GENE MUTATIONS AND
;; FILE REFERENCE: 13761-706US1
;; CURRENT APPLICATION NUMBER: US/09/235.538
;; CURRENT FILING DATE: 1999-01-22
;; PRIOR APPLICATION NUMBER: US 60/072,225
;; PRIOR FILING DATE: 1998-01-23
;; PRIOR APPLICATION NUMBER: PCT/US99/01165
;; PRIOR FILING DATE: 1999-01-20
;; NUMBER OF SEQ ID NOS: 7
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1
;; LENGTH: 2437
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;;
US-09-235-538-1

Query Match 2.7%; Score 31; DB 4; Length 2437;
Best Local Similarity 100.0%; Pred. No. 1.le-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1012 TATATACATATATATATATATATATATAT 1042
|||||
Db 926 TATATACATATATATATATATATATATAT 896

Search completed: July 17, 2003, 17:05:46
Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 16:36:32 ; Search time 311 Seconds
(without alignments)
7681.527 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 1158

Sequence: 1 aatgagcaaccgtgtcaaca.....cctctctgatcaatgcttt 1158

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1439767 seqs, 1031500376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
- 11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1158	100.0	1158	11	US-09-814-661A-1
2	45	3.9	65	11	US-09-814-661A-5
3	45	3.9	65	11	US-09-814-661A-6
4	33	2.8	1018	15	US-10-027-632-262771
5	32	2.8	448	15	US-10-027-632-36437
6	32	2.8	448	15	US-10-027-632-361325
7	32	2.8	448	15	US-10-027-632-298177
8	32	2.8	642	15	US-10-027-632-286954
9	32	2.8	642	15	US-10-027-632-286956
10	32	2.8	672	15	US-10-027-632-110418
11	32	2.8	672	15	US-10-027-632-110419
12	32	2.8	672	15	US-10-027-632-110420
13	31	2.7	40	15	US-10-085-906-177
14	31	2.7	162	11	US-09-969-373-733
15	31	2.7	200	12	US-09-754-853A-386
16	31	2.7	348	15	US-10-027-632-285347

ALIGNMENTS

RESULT 1
US-09-814-661A-1
; Sequence 1, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814, 661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: S. Cerevisiae
US-09-814-661A-1

Query Match 100.0%; Score 1158; DB 11; Length 1158;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AATGAGCAACCGTGTCAACAGAGTGTCAAGACCGGTACTATTATCCCAAGGATACG	60
Db	1	AATGAGCAACCGTGTCAACAGAGTGTCAAGACCGGTACTATTATCCCAAGGATACG	60
Qy	61	TCCTTCTGCGCAACATCATTCGCGTCCGAACGTCGCGGGGTCCTTCTGACATTCG	120
Db	61	TCCTTCTGCGCAACATCATTCGCGTCCGAACGTCGCGGGGTCCTTCTGACATTCG	120
Qy	121	TACTTCCCACTAAGAGCATCTTCTCTCTTTTGTAGGCCAATGATAGGAAGACAA	180
Db	121	TACTTCCCACTAAGAGCATCTTCTCTCTTTTGTAGGCCAATGATAGGAAGACAA	180
Qy	181	TAGATTATAAATACGTCAGAAATATAGTAGATATGTTTTATGTTTACCTCGTACAT	240
Db	181	TAGATTATAAATACGTCAGAAATATAGTAGATATGTTTTATGTTTACCTCGTACAT	240

```

QY 241 GAATATTACGCTTTTTCCTGCGCAACATTTGAAATTTTTTTTGTGTTACCTCGCGCTGA 300
Db 241 GAATATTACGCTTTTTCCTGCGCAACATTTGAAATTTTTTTTGTGTTACCTCGCGCTGA 300
QY 301 GCCAAACGGGTCCACTACCGCGCGCGCTGCGCAATTTTGGGAAGTCACTCCGTCGCAAAA 360
Db 301 GCCAAACGGGTCCACTACCGCGCGCGCTGCGCAATTTTGGGAAGTCACTCCGTCGCAAAA 360
QY 361 AGGAATAGCCATACATATGTTTACTGTTTGGGAACATCGCGCGGTTCGCCCGGATTCGG 420
Db 361 AGGAATAGCCATACATATGTTTACTGTTTGGGAACATCGCGCGGTTCGCCCGGATTCGG 420
QY 421 CTTACGCGGTATAAAGAGATCTTTTTCCTGCGCTGCTCCCTCCATTTTAAATG 480
Db 421 CTTACGCGGTATAAAGAGATCTTTTTCCTGCGCTGCTCCCTCCATTTTAAATG 480
QY 481 TCTATCTGCTCTTGTGATCTTACGCTCTCACTAACCTCTCTCAACTGCTCAATAAT 540
Db 481 TCTATCTGCTCTTGTGATCTTACGCTCTCACTAACCTCTCTCAACTGCTCAATAAT 540
QY 541 TTCCCGCTATGAAATTTCCCAAGACTACTTTTACGCTCAAAATCGCTGCGCAACACAC 600
Db 541 TTCCCGCTATGAAATTTCCCAAGACTACTTTTACGCTCAAAATCGCTGCGCAACACAC 600
QY 601 AAGCCCTTCCCATTTGCGTACCGTACCATGCGGAATTTAGAGGGTGCCTTTGCCAC 660
Db 601 AAGCCCTTCCCATTTGCGTACCGTACCATGCGGAATTTAGAGGGTGCCTTTGCCAC 660
QY 661 CTATGCTGAGGTCTCTATGTTGTCTACTCAAACTCCATGGCAGCTCCGCTTCTGCT 720
Db 661 CTATGCTGAGGTCTCTATGTTGTCTACTCAAACTCCATGGCAGCTCCGCTTCTGCT 720
QY 721 CGGCTCTTCATAGAAATGTTGGGAAGGATTTGGAGGAGACTCAACTCTATCGATC 780
Db 721 CGGCTCTTCATAGAAATGTTGGGAAGGATTTGGAGGAGACTCAACTCTATCGATC 780
QY 781 ATGACATGAACACAAATTTGTTCTGGGAACTAAATCTATGTTCAACCGAGGTA 840
Db 781 ATGACATGAACACAAATTTGTTCTGGGAACTAAATCTATGTTCAACCGAGGTA 840
QY 841 AGGTCAGGAATGAGTCTTAAAGTTCCTTTCTACTACTCTTTCTTTCTTTCTTTCTTT 900
Db 841 AGGTCAGGAATGAGTCTTAAAGTTCCTTTCTACTACTCTTTCTTTCTTTCTTTCTTT 900
QY 901 CCACATAGTCTGTTCTTTCTCTCTAGATACCTCTTTTCAGGGACTCTCGTCTTA 960
Db 901 CCACATAGTCTGTTCTTTCTCTCTAGATACCTCTTTTCAGGGACTCTCGTCTTA 960
QY 961 CTATGTTGTCATTTCTGAAACATTTCTCTCGGTGCAATTTCTTTCTTTCTTTATATACAT 1020
Db 961 CTATGTTGTCATTTCTGAAACATTTCTCTCGGTGCAATTTCTTTCTTTCTTTATATACAT 1020
QY 1021 ATATATATATATATATATATGCTCTCTCTACGATTTTGTATTTCTGCTCTTTATC 1080
Db 1021 ATATATATATATATATATATGCTCTCTCTACGATTTTGTATTTCTGCTCTTTATC 1080
QY 1081 AAGATAGTCTATATACCTTTTCATACAGCTAGATATCCCTAGCGCCACATTTGCCCC 1140
Db 1081 AAGATAGTCTATATACCTTTTCATACAGCTAGATATCCCTAGCGCCACATTTGCCCC 1140
QY 1141 TCTCTTGATCAATGCTTT 1158
Db 1141 TCTCTTGATCAATGCTTT 1158

```

RESULT 2

```

US-09-814-661A-5
; Sequence 5, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan

```

```

; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814,661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer
US-09-814-661A-5

```

```

Query Match 3.9%; Score 45; DB 11; Length 65;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 502 CTTACGGTCTCACTAACCTCTCTTCAACTGCTCAATAATTTCCCG 546
Db 1 CTTACGGTCTCACTAACCTCTCTTCAACTGCTCAATAATTTCCCG 45

```

RESULT 3

```

US-09-814-661A-6/C
; Sequence 6, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814,661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer
US-09-814-661A-6

```

```

Query Match 3.9%; Score 45; DB 11; Length 65;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 833 CCAGGTAGGTCGAGGAATGAGTCTTAAAGTTCCTTTCATAC 877
Db 45 CCAGGTAGGTCGAGGAATGAGTCTTAAAGTTCCTTTCATAC 1

```

RESULT 4

```

US-10-027-632-262771
; Sequence 262771, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 10827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24

```



```

; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 262771
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-262771

```

```

Query Match          2.8%; Score 33; DB 15; Length 1018;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

-Qy 1010 TTTATATACATATATATATATATATATATATAT 1042
      |||||||||||||||||||||||||||||||||||
Db 111 TTTATATACATATATATATATATATATATATAT 143

```

```

RESULT 5
US-10-027-632-36437/c
; Sequence 36437, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36437
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-36437

```

```

Query Match          2.8%; Score 32; DB 15; Length 448;
Best Local Similarity 100.0%; Pred. No. 4.8e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

-Qy 1015 ATACATATATATATATATATATATATATGTCT 1046
      |||||||||||||||||||||||||||||||||||
Db 421 ATACATATATATATATATATATATATATGTCT 390

```

```

RESULT 6
US-10-027-632-61325/c
; Sequence 61325, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632

```

```

; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61325
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-61325

```

```

Query Match          2.8%; Score 32; DB 15; Length 448;
Best Local Similarity 100.0%; Pred. No. 4.8e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

-Qy 1015 ATACATATATATATATATATATATATGTCT 1046
      |||||||||||||||||||||||||||||||||||
Db 421 ATACATATATATATATATATATATATGTCT 390

```

```

RESULT 7
US-10-027-632-298177/c
; Sequence 298177, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 298177
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-298177

```

```

Query Match          2.8%; Score 32; DB 15; Length 448;
Best Local Similarity 100.0%; Pred. No. 4.8e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

-Qy 1015 ATACATATATATATATATATATATATGTCT 1046
      |||||||||||||||||||||||||||||||||||
Db 421 ATACATATATATATATATATATATATGTCT 390

```

RESULT 8
US-10-027-632-286954/c
; Sequence 286954, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 286954
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-286954

Query Match 2.8%; Score 32; DB 15; Length 642;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1012 TATATACATATATATATATATATATG 1043
Db 344 TATATACATATATATATATATATATG 313

RESULT 9
US-10-027-632-286956/c
; Sequence 286956, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 286956
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-286956

Query Match 2.8%; Score 32; DB 15; Length 642;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1012 TATATACATATATATATATATATATG 1043
Db 344 TATATACATATATATATATATATATG 313

RESULT 10
US-10-027-632-110418
; Sequence 110418, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110418
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-110418

Query Match 2.8%; Score 32; DB 15; Length 672;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1012 TATATACATATATATATATATATATG 1043
Db 334 TATATACATATATATATATATATATG 365

RESULT 11
US-10-027-632-110419
; Sequence 110419, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110419
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-110419

Query Match 2.8%; Score 32; DB 15; Length 672;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATATG 1043
|||||
Db 334 TATATACATATATATATATATATATATATG 365

RESULT 12
US-10-027-632-110420
; Sequence 110420, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110420
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-110420

Query Match 2.8%; Score 32; DB 15; Length 672;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATATG 1043
|||||
Db 334 TATATACATATATATATATATATATATATG 365

RESULT 13
US-10-085-906-177
; Sequence 177, Application US/10085906
; Publication No. US20030054371A1
; GENERAL INFORMATION:
; APPLICANT: Ying, Vincent
; APPLICANT: Wu, Paul
; APPLICANT: Gray, Gary S.
; TITLE OF INVENTION: POLYMORPHIC ELEMENTS IN THE
; TITLE OF INVENTION: COSTIMULATORY RECEPTOR LOCUS AND USES THEREOF
; FILE REFERENCE: GNN-5343CP2
; CURRENT APPLICATION NUMBER: US/10/085,906
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 60/126,215

; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 09/534,061
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: PCT/US00/07938
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 545
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 177
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-906-177

Query Match 2.7%; Score 31; DB 15; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATAT 1042
|||||
Db 2 TATATACATATATATATATATATATATATATAT 32

RESULT 14
US-09-969-373-733/c
; Sequence 733, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 733
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-733

Query Match 2.7%; Score 31; DB 11; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1012 TATATACATATATATATATATATATATAT 1042
|||||
Db 116 TATATACATATATATATATATATATATATAT 86

RESULT 15
US-09-754-853A-386/c
; Sequence 386, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 386

```

: LENGTH: 200
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: 515002_region_G2_63275_46
US-09-754-853A-386

Query Match      2.7%  Score 31;  DB 12; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy  1012 TATATACATATATATATATATATATATATAT 1042
      |||||||
Db   172 TATATACATATATATATATATATATATATAT 142

Search completed: July 17, 2003, 18:10:24
Job time : 312 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 19:59:51 ; Search time 2239 Seconds
(without alignments)
1167.851 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 104
Sequence: 1 MQNSQDYFYAQNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Word size: 1

Total number of hits satisfying chosen parameters: 49569006

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09814661/runat_16072003_115402_15665/app_query.fasta_1.263
-DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=oligo -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661 -CGN_1_1_4447 -runat_16072003_115402_15665 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Pending_Patents_NA_Main :*

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq :*

2: /cgn2_6/ptodata/1/pna/US06_COMB.seq :*

3: /cgn2_6/ptodata/1/pna/US07_COMB.seq :*

4: /cgn2_6/ptodata/1/pna/US080_COMB.seq :*

5: /cgn2_6/ptodata/1/pna/US081_COMB.seq :*

6: /cgn2_6/ptodata/1/pna/US082_COMB.seq :*

7: /cgn2_6/ptodata/1/pna/US083_COMB.seq :*

8: /cgn2_6/ptodata/1/pna/US084_COMB.seq :*

9: /cgn2_6/ptodata/1/pna/US085_COMB.seq :*

10: /cgn2_6/ptodata/1/pna/US086_COMB.seq :*

11: /cgn2_6/ptodata/1/pna/US087_COMB.seq :*

12: /cgn2_6/ptodata/1/pna/US088_COMB.seq :*

13: /cgn2_6/ptodata/1/pna/US089_COMB.seq :*

14: /cgn2_6/ptodata/1/pna/US090_COMB.seq :*

15: /cgn2_6/ptodata/1/pna/US091_COMB.seq :*

16: /cgn2_6/ptodata/1/pna/US092_COMB.seq :*

17: /cgn2_6/ptodata/1/pna/US093_COMB.seq :*

18: /cgn2_6/ptodata/1/pna/US094_COMB.seq :*

19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq :*

20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq :*

21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq :*

22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq :*

23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq :*

24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq :*

25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq :*

26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq :*

27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq :*

28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq :*

29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq :*

30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq :*

31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq :*

32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq :*

33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq :*

34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq :*

35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq :*

36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq :*

37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq :*

38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq :*

39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq :*

40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq :*

41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq :*

42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq :*

43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq :*

44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq :*

45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq :*

46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq :*

47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq :*

48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq :*

49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq :*

50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq :*

51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq :*

52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq :*

53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq :*

54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq :*

55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq :*

56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq :*

57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq :*

58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq :*

59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq :*

60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq :*

61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq :*

62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq :*

63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq :*

64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq :*

65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq :*

66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq :*

67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq :*

68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq :*

69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq :*

70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq :*

71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq :*

72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq :*

73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq :*

74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq :*

75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq :*

76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq :*

77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq :*

78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq :*

79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq :*

80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq :*

81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq :*

82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq :*

83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq :*

84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq :*

85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq :*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	1158	15	US-09-159-858-1
2	104	100.0	1158	31	US-09-814-661A-1
3	104	100.0	924430	14	US-09-012-031-314
4	104	100.0	924430	14	US-09-012-031B-314
5	104	100.0	924430	14	US-09-012-031C-314
6	104	100.0	924430	17	US-09-335-032-12216

Sequence 1, Appli
Sequence 1, Appli
Sequence 314, App
Sequence 314, App
Sequence 314, App

```
C 7 10 9.6 254 17 US-09-304-517A-35297 Sequence 35297, A
C 8 10 9.6 254 17 US-09-371-146A-35297 Sequence 35297, A
C 9 10 9.6 254 18 US-09-436-711-5436 Sequence 5426, A
C 10 10 9.6 254 37 US-09-985-678-35297 Sequence 35297, A
C 11 10 9.6 254 38 US-10-035-416-5436 Sequence 5426, A
C 12 10 9.6 301 17 US-09-303-031A-7828 Sequence 7828, Ap
C 13 10 9.6 301 33 US-09-894-949A-7828 Sequence 7828, Ap
C 14 10 9.6 301 33 US-09-894-949A-7828 Sequence 7828, Ap
C 15 10 9.6 302 17 US-09-304-517A-58806 Sequence 58806, A
C 16 10 9.6 302 17 US-09-371-146A-58806 Sequence 58806, A
C 17 10 9.6 302 37 US-09-985-678-58806 Sequence 58806, A
C 18 10 9.6 316 17 US-09-304-517A-59136 Sequence 59136, A
C 19 10 9.6 316 17 US-09-371-146A-59136 Sequence 59136, A
C 20 10 9.6 316 37 US-09-985-678-59136 Sequence 59136, A
C 21 10 9.6 318 17 US-09-303-031A-8484 Sequence 8484, Ap
C 22 10 9.6 318 33 US-09-894-949A-8484 Sequence 8484, Ap
C 23 10 9.6 318 33 US-09-894-949A-8484 Sequence 8484, Ap
C 24 10 9.6 384 63 US-09-197-872-42793 Sequence 42793, A
C 25 10 9.6 446 18 US-09-421-106-36229 Sequence 36229, A
C 26 10 9.6 466 28 US-09-705-926-14661 Sequence 14661, A
C 27 10 9.6 545 25 US-09-654-617-281012 Sequence 281012,
C 28 10 9.6 545 27 US-09-684-016-281012 Sequence 281012,
C 29 10 9.6 577 43 US-10-260-238-390 Sequence 390, App
C 30 10 9.6 577 76 US-09-325-448-390 Sequence 390, App
C 31 10 9.6 578 25 US-09-654-617-397003 Sequence 397003,
C 32 10 9.6 578 27 US-09-684-016-397003 Sequence 397003,
C 33 10 9.6 578 29 US-09-733-089-10027 Sequence 10027, A
C 34 10 9.6 578 31 US-09-816-660-10027 Sequence 10027, A
C 35 10 9.6 634 33 US-09-873-402A-6964 Sequence 6964, A
C 36 10 9.6 634 64 US-09-209-830-6964 Sequence 6964, A
C 37 10 9.6 1014 41 US-10-155-881-24308 Sequence 24308, A
C 38 10 9.6 1426 28 US-09-705-926-1888 Sequence 1888, Ap
C 39 10 9.6 1482 28 US-09-708-427-67775 Sequence 67775, A
C 40 10 9.6 2001 16 US-09-270-767-31380 Sequence 31380, A
C 41 10 9.6 2001 16 US-09-270-767-31380 Sequence 31380, A
C 42 10 9.6 2165 16 US-09-270-767-15080 Sequence 15080, A
C 43 10 9.6 2542 80 US-09-360-207-18401 Sequence 18401, A
C 44 10 9.6 2616 80 US-09-360-039-42226 Sequence 42226, A
C 45 10 9.6 3423 23 US-09-614-150-33986 Sequence 33986, A
```

ALIGNMENTS

```
RESULT 1
US-09-159-858-1
; Sequence 1, Application US/09159858
; GENERAL INFORMATION:
```

```
; APPLICANT: Rothstein, Rodney
```

```
; APPLICANT: Zhao, Xiaolan
```

```
; TITLE OF INVENTION: A Small Protein That Interacts With A Ribonucleotide
```

```
; FILE REFERENCE: 56615
```

```
; CURRENT APPLICATION NUMBER: US/09/159,858
```

```
; CURRENT FILING DATE: 1998-09-24
```

```
; NUMBER OF SEQ ID NOS: 22
```

```
; SOFTWARE: PatentIn ver. 2.1
```

```
; SEQ ID NO 1
```

```
; LENGTH: 1158
```

```
; TYPE: DNA
```

```
; ORGANISM: S. cerevisiae
```

```
US-09-159-858-1
```

```
Alignment Scores:
Pred. No.: 1.45e-84 Length: 1158
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0
```

```
US-09-814-661A-2 (1-104) x US-09-159-858-1 (1-1158)
```

```
Qy 1 MetClnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
```

```
Db 549 ATGCAAAATCCCAAGACTACTTTTACGCTCAAAATCGCTCCCAACAACAAGCCCT 608
Qy 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPtoMetAla 40
Db 609 TCCACATTCGCTACCGTGACCATGGCGGAATTTAGAAAGGTGGCTTTGCCACCTATGGCT 668
Qy 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
Db 669 GAGGTTCTCTATGTTGCTACTCAAAACTCCATGGCAGCTCCGCTTCTGCCTCCGCTTCT 728
Qy 61 SerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAspHisAspMet 80
Db 729 TCATTAGAAATGCGGAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 788
Qy 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
Db 789 AACAAACAACAATTTGGTCTTGGCGAACTAAATCTATGTTCAACCAAGGTAAGTCCGAG 848
Qy 101 GluMetAspPhe 104
Db 849 GAAATGGACTTC 860
RESULT 2
US-09-814-661A-1
; Sequence 1, Application US/09814661A
; GENERAL INFORMATION:
```

```
; APPLICANT: Rothstein, Rodney
```

```
; APPLICANT: Zhao, Xiaolan
```

```
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
```

```
; FILE REFERENCE: 0575/56615-A-PCT-US
```

```
; CURRENT APPLICATION NUMBER: US/09/814,661A
```

```
; CURRENT FILING DATE: 2001-03-22
```

```
; NUMBER OF SEQ ID NOS: 27
```

```
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 1
```

```
; LENGTH: 1158
```

```
; TYPE: DNA
```

```
; ORGANISM: S. Cerevisiae
```

```
US-09-814-661A-1
```

```
Alignment Scores:
Pred. No.: 1.45e-84 Length: 1158
Score: 104.00 Matches: 104
```

```
Percent Similarity: 100.00% Conservative: 0
```

```
Best Local Similarity: 100.00% Mismatches: 0
```

```
Query Match: 100.00% Indels: 0
```

```
DB: 31 Gaps: 0
```

```
US-09-814-661A-2 (1-104) x US-09-814-661A-1 (1-1158)
```

```
Qy 1 MetClnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
```

```
Db 549 ATGCAAAATCCCAAGACTACTTTTACGCTCAAAATCGCTCCCAACAACAAGCCCT 608
```

```
Qy 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPtoMetAla 40
```

```
Db 609 TCCACATTCGCTACCGTGACCATGGCGGAATTTAGAAAGGTGGCTTTGCCACCTATGGCT 668
```

```
Qy 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
```

```
Db 669 GAGGTTCTCTATGTTGCTACTCAAAACTCCATGGCAGCTCCGCTTCTGCCTCCGCTTCT 728
```

```
Qy 61 SerLeuGluMetTrpGluLysAspLeuGluArgLeuAsnSerIleAspHisAspMet 80
```

```
Db 729 TCATTAGAAATGCGGAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 788
```

```
Qy 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
```

```
Db 789 AACAAACAACAATTTGGTCTTGGCGAACTAAATCTATGTTCAACCAAGGTAAGTCCGAG 848
```

```
Qy 101 GluMetAspPhe 104
```

```

; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031B-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031B-314 (1-924430)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
DB 159383 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGCTCCAAACAACAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
DB 159443 TCCACATTCGCTACCGTGACCATGGCGAATTTAGAGGGTGCCCTTCCACCTATGGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
DB 159503 GAGGTTCCCTATGTTGCTACTCAAAACTCCATGGCGAGCTCGGCTTCTCGCTTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
DB 159563 TCATTAGAAATGTGGGAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGCATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
DB 159623 AACACAAACAATTTGGTTCTGGCGAACTAAATCTATGTTCAACAGGGTAAGTGCAG 159682
QY 101 GluMetAspPhe 104
DB 159683 GAAATGGACTTC 159694

RESULT 5
US-09-012-031C-314
; Sequence 314, Application US/09012031C
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 1107.73251
; CURRENT APPLICATION NUMBER: US/09/012,031C
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/035,917
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031C-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031C-314 (1-924430)

```

```

; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031B-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031B-314 (1-924430)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 159383 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGTCCAAACAACAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
Db 159443 TCCACATTCGCTACCGTGACCATGGCGGAATTTAGAGGGTGCCCTTCCACCATGATGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
Db 159503 GAGGTTCCCTATGTTGCTACTCAAAACTCCATGGCGAGCTCGGCTTCGCTCCGCTTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
Db 159563 TCATTAGAAATGTGGGAAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
Db 159623 AACACAAACAATTTGGTTCTGGCGAACTAAATCTATGTTCAACAGGGTAAGTGCAG 159682
QY 101 GluMetAspPhe 104
Db 159683 GAAATGGACTTC 159694

RESULT 5
US-09-012-031C-314
; Sequence 314, Application US/09012031C
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 1107.73251
; CURRENT APPLICATION NUMBER: US/09/012,031C
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/035,917
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031C-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031C-314 (1-924430)

; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031B-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031B-314 (1-924430)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 159383 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGTCCAAACAACAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
Db 159443 TCCACATTCGCTACCGTGACCATGGCGGAATTTAGAGGGTGCCCTTCCACCATGATGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
Db 159503 GAGGTTCCCTATGTTGCTACTCAAAACTCCATGGCGAGCTCGGCTTCGCTCCGCTTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
Db 159563 TCATTAGAAATGTGGGAAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
Db 159623 AACACAAACAATTTGGTTCTGGCGAACTAAATCTATGTTCAACAGGGTAAGTGCAG 159682
QY 101 GluMetAspPhe 104
Db 159683 GAAATGGACTTC 159694

RESULT 4
US-09-012-031B-314
; Sequence 314, Application US/09012031B
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 1107.73251
; CURRENT APPLICATION NUMBER: US/09/012,031B
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 60/035,917
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031B-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031B-314 (1-924430)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 159383 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGTCCAAACAACAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
Db 159443 TCCACATTCGCTACCGTGACCATGGCGGAATTTAGAGGGTGCCCTTCCACCATGATGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
Db 159503 GAGGTTCCCTATGTTGCTACTCAAAACTCCATGGCGAGCTCGGCTTCGCTCCGCTTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
Db 159563 TCATTAGAAATGTGGGAAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
Db 159623 AACACAAACAATTTGGTTCTGGCGAACTAAATCTATGTTCAACAGGGTAAGTGCAG 159682
QY 101 GluMetAspPhe 104
Db 159683 GAAATGGACTTC 159694

RESULT 3
US-09-012-031A-314
; Sequence 314, Application US/09012031A
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 1107.73251
; CURRENT APPLICATION NUMBER: US/09/012,031A
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 60/035,917
; NUMBER OF SEQ ID NOS: 317
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031A-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031A-314 (1-924430)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 159383 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGTCCAAACAACAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
Db 159443 TCCACATTCGCTACCGTGACCATGGCGGAATTTAGAGGGTGCCCTTCCACCATGATGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
Db 159503 GAGGTTCCCTATGTTGCTACTCAAAACTCCATGGCGAGCTCGGCTTCGCTCCGCTTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
Db 159563 TCATTAGAAATGTGGGAAAAGGATTTGGAGGAGAGACTCAACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
Db 159623 AACACAAACAATTTGGTTCTGGCGAACTAAATCTATGTTCAACAGGGTAAGTGCAG 159682
QY 101 GluMetAspPhe 104
Db 159683 GAAATGGACTTC 159694

RESULT 2
US-09-012-031B-314
; Sequence 314, Application US/09012031B
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 1107.73251
; CURRENT APPLICATION NUMBER: US/09/012,031B
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 60/035,917
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-012-031B-314

Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-012-031B-314 (1-924430)

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
Db 159383 ATGCAAAATTCCTCAAGACTACTTTTACGCTCAAAATCGTCCAAACAACAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla
```

QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
159383 ATGCAAAATTCACAGACTACTTTACGCTCAAAATCGCTGCCACACACAAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
159443 TCCACATTGCGTACCGTGACCATGGCGGAATTTAGAGGGTGCCTTTGCCACCTATGGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSer 60
159503 GAGGTTCTCTATGTTCTACTCAAACTCCATGGGCGAGCTCGCTTCTGCCCTCCGCTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
159563 TCATTAGAAATCTGGGAAAGGATTTGGAGGAGACACTCACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
159623 AACACACAAATTTGGTTCTGCGAACTAAATCTATGTTCAACGAGGTAAAGTCGAG 159682
QY 101 GluMetAspPhe 104
159683 GAAATGGACTTC 159694
Db 159683 GAAATGGACTTC 159694
RESULT 6
US-09-335-032-12216
; Sequence 12216, Application US/09335032
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 01107.78572
; CURRENT APPLICATION NUMBER: US/09/335,032
; PRIOR FILING DATE: 1999-06-16
; PRIOR APPLICATION NUMBER: US 60/035,917
; PRIOR FILING DATE: 1997-01-23
; PRIOR APPLICATION NUMBER: US 09/012,031
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 12219
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12216
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-335-032-12216
Alignment Scores:
Pred. No.: 4,23e-82 Length: 924430
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0
US-09-814-661A-2 (1-104) x US-09-335-032-12216 (1-924430)
QY 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
159383 ATGCAAAATTCACAGACTACTTTACGCTCAAAATCGCTGCCACACACAAAGCCCT 159442
QY 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProMetAla 40
159443 TCCACATTGCGTACCGTGACCATGGCGGAATTTAGAGGGTGCCTTTGCCACCTATGGCT 159502
QY 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSer 60
159503 GAGGTTCTCTATGTTCTACTCAAACTCCATGGGCGAGCTCGCTTCTGCCCTCCGCTCT 159562
QY 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
159563 TCATTAGAAATCTGGGAAAGGATTTGGAGGAGACACTCACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
159623 AACACACAAATTTGGTTCTGCGAACTAAATCTATGTTCAACGAGGTAAAGTCGAG 159682
QY 101 GluMetAspPhe 104
159683 GAAATGGACTTC 159694
Db 159683 GAAATGGACTTC 159694

Db 159563 TCATTAGAAATCTGGGAAAGGATTTGGAGGAGAGACTCACTCTATCGATCATGACATG 159622
QY 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
159623 AACACACAAATTTGGTTCTGCGGCAACTAAATCTATGTTCAACGAGGTAAAGTCGAG 159682
QY 101 GluMetAspPhe 104
159683 GAAATGGACTTC 159694
Db 159683 GAAATGGACTTC 159694
RESULT 7
US-09-304-517A-35297/c
; Sequence 35297, Application US/09304517A
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)B
; CURRENT APPLICATION NUMBER: US/09/304,517A
; CURRENT FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 35297
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Zea mays
US-09-304-517A-35297
Alignment Scores:
Pred. No.: 16.6 Length: 254
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 17 Gaps: 0
US-09-814-661A-2 (1-104) x US-09-304-517A-35297 (1-254)
QY 53 SerSerAlaSerAlaSerAlaSerSerLeu 62
151 TCCCTCGCTTCTGCTTCTGCTTCTGCTTCTTCTCTT 122
Db 151 TCCCTCGCTTCTGCTTCTGCTTCTGCTTCTTCTCTT 122
RESULT 8
US-09-371-146A-35297/c
; Sequence 35297, Application US/09371146A
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: ANNOTATED PLANT GENES
; FILE REFERENCE: 38-21(15097)C
; CURRENT APPLICATION NUMBER: US/09/371,146A
; CURRENT FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 294310
; SEQ ID NO 35297
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Zea mays
US-09-371-146A-35297
Alignment Scores:
Pred. No.: 16.6 Length: 254
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 17 Gaps: 0
US-09-814-661A-2 (1-104) x US-09-371-146A-35297 (1-254)
QY 53 SerSerAlaSerAlaSerAlaSerSerLeu 62
151 TCCCTCGCTTCTGCTTCTGCTTCTGCTTCTTCTCTT 122
Db 151 TCCCTCGCTTCTGCTTCTGCTTCTGCTTCTTCTCTT 122


```
RESULT 9
US-09-436-711-5426/c
; Sequence 5426, Application US/09436711
; GENERAL INFORMATION:
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Gooding, Douglas, H.
; APPLICANT: Sherman, Bradley, K.
; APPLICANT: Wu, Kunsheng
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN LEAF
; FILE REFERENCE: PL-0004 US
; CURRENT APPLICATION NUMBER: US/09/436,711
; CURRENT FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: 60/077,830
; PRIOR FILING DATE: March 13, 1998
; NUMBER OF SEQ ID NOS: 6518
; SOFTWARE: PERL Program
; SEQ ID NO 5426
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 700223279H1
US-09-436-711-5426

Alignment Scores:
Pred. No.: 16.6 Length: 254
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 18 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-436-711-5426 (1-254)
QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 151 TCCTCGCTTCTGCTTCTGCTTCTTCTT 122

RESULT 10
US-09-985-678-35297/c
; Sequence 35297, Application US/09985678
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordin
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US/09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 35297
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Zea mays
US-09-985-678-35297

Alignment Scores:
Pred. No.: 16.6 Length: 254
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 37 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-985-678-35297 (1-254)
QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 151 TCCTCGCTTCTGCTTCTGCTTCTTCTT 122

RESULT 11
US-10-035-416-5426/c
; Sequence 5426, Application US/10035416
; GENERAL INFORMATION:
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Gooding, Douglas, H.
; APPLICANT: Sherman, Bradley, K.
; APPLICANT: Wu, Kunsheng
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN LEAF
; FILE REFERENCE: PL-0004 US
; CURRENT APPLICATION NUMBER: US/10/035,416
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 09/436,711
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 6518
; SOFTWARE: PERL Program
; SEQ ID NO 5426
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 700223279H1
US-10-035-416-5426

Alignment Scores:
Pred. No.: 16.6 Length: 254
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 38 Gaps: 0

US-09-814-661A-2 (1-104) x US-10-035-416-5426 (1-254)
QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 151 TCCTCGCTTCTGCTTCTGCTTCTTCTT 122

RESULT 12
US-09-303-031A-7828/c
; Sequence 7828, Application US/09303031A
; GENERAL INFORMATION:
; APPLICANT: Raghunath, Lalagudi
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN CALLUS
; FILE REFERENCE: PL-0011 US
; CURRENT APPLICATION NUMBER: US/09/303,031A
; CURRENT FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/084,493
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 9627
; SOFTWARE: PERL Program
; SEQ ID NO 7828
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 700474247H1
US-09-303-031A-7828

Alignment Scores:
Pred. No.: 19.1 Length: 301
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-303-031A-7828 (1-301)
```

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 161 TCCTCTGCTTCGCTTCGCTTCCTCTCT 132

RESULT 13

US-09-894-949-7828/c
; Sequence 7828, Application US/09894949
; GENERAL INFORMATION:
; APPLICANT: Lalguadi, Raghunath V.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN CALLUS
; FILE REFERENCE: PL-0011-1 CON
; CURRENT APPLICATION NUMBER: US/09/894,949
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/303,031
; PRIOR FILING DATE: April 29, 1999
; PRIOR APPLICATION NUMBER: 60/084,493
; PRIOR FILING DATE: May 6, 1998
; NUMBER OF SEQ ID NOS: 9627
; SOFTWARE: PERL Program
; SEQ ID NO 7828
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 700474247H1
US-09-894-949-7828

Alignment Scores:
Pred. No.: 19.1 Length: 301
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 33 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-894-949-7828 (1-301)

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 161 TCCTCTGCTTCGCTTCGCTTCCTCTCT 132

RESULT 14

US-09-894-949A-7828/c
; Sequence 7828, Application US/09894949A
; GENERAL INFORMATION:
; APPLICANT: Raghunath, Lalguadi
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN CALLUS
; FILE REFERENCE: PL-0011 US
; CURRENT APPLICATION NUMBER: US/09/894,949A
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/303,031
; PRIOR FILING DATE: 1999-04-29
; NUMBER OF SEQ ID NOS: 9627
; SOFTWARE: PERL Program
; SEQ ID NO 7828
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 700474247H1
US-09-894-949A-7828

Alignment Scores:
Pred. No.: 19.1 Length: 301

Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 33 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-894-949A-7828 (1-301)

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 161 TCCTCTGCTTCGCTTCGCTTCCTCTCT 132

RESULT 15

US-09-304-517A-58806/c
; Sequence 58806, Application US/09304517A
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)B
; CURRENT APPLICATION NUMBER: US/09/304,517A
; CURRENT FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 58806
; LENGTH: 302
; TYPE: DNA
; ORGANISM: Zea mays
US-09-304-517A-58806

Alignment Scores:
Pred. No.: 19.2 Length: 302
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 17 Gaps: 0

US-09-814-661A-2 (1-104) x US-09-304-517A-58806 (1-302)

QY 53 SerSerAlaSerAlaSerAlaSerLeu 62
|||||
Db 161 TCCTCTGCTTCGCTTCGCTTCCTCTCT 132

Search completed: July 17, 2003, 21:32:08
Job time : 2582 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 12:02:20 ; Search time 56 Seconds
(without alignments)

5510.865 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 375

Sequence: 1 aatgagcaacccgtgtaaca.....cctctcttgatcaatgcttt 1158

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0.

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09814661/runat.16072003.115330.15166/app.query.fasta.1.1351
-DB=A.Geneseq.101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661 -CGN.1.114 -runat.16072003.115330.15166 -NCPU=3
-NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A.Geneseq.101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	104	27.7	104	21	AA90470	Yeast Sm11 (suppre
2	11	2.9	2206	21	AAB18254	Plasmodium falci
3	10	2.7	19	23	AAM51850	P falciparum isop
4	10	2.7	25	21	AAB33773	Human secreted pr
5	10	2.7	35	21	AAB34410	Gene 34 human secr
6	10	2.7	47	21	AAB34207	Gene 43 human secr
7	10	2.7	48	23	AAM51856	P falciparum isop
8	10	2.7	62	22	AAO08414	Human polypeptide
9	10	2.7	62	22	AAO12066	Human polypeptide
10	10	2.7	67	22	AAO11381	Human polypeptide
11	10	2.7	67	22	AAO12482	Human polypeptide
12	10	2.7	69	22	AAO11077	Human polypeptide
13	10	2.7	73	23	ABP42283	Human ovarian anti
14	10	2.7	75	22	AAU31237	Novel human secret
15	10	2.7	78	22	AAO00992	Human polypeptide
16	10	2.7	78	22	AAO11510	Human polypeptide
17	10	2.7	84	22	AAO07427	Human polypeptide
18	10	2.7	86	22	AAO75549	Human colon cancer
19	10	2.7	89	22	AAO12651	Human polypeptide
20	10	2.7	96	22	AAO11694	Human polypeptide
21	10	2.7	100	22	AAO07396	Human polypeptide
22	10	2.7	138	22	AAO00532	Human polypeptide
23	10	2.7	238	21	AAB34720	Human secreted pro
24	10	2.7	269	21	AAB18237	Plasmodium falci
25	10	2.7	355	22	ABG00422	Novel human diagno
26	10	2.7	362	17	AAW02082	Novel human diagno
27	10	2.7	463	20	AAO13392	Amino acid sequenc
28	10	2.7	463	21	AAB01373	Neuron-associated
29	10	2.7	463	21	AAO95343	Human PRO328 antit
30	10	2.7	463	22	AAU12351	Human PRO328 poly
31	10	2.7	463	22	AAB88408	Human membrane or
32	10	2.7	463	22	AAB80260	Human PRO328 prote
33	10	2.7	463	22	AAO53088	Human angioogenesis
34	10	2.7	1186	20	AAO58399	Banana ripening fr
35	10	2.7	1186	20	AAO58399	Banana ripening fr
36	10	2.7	1187	20	AAO58401	Banana ripening fr
37	10	2.7	3096	20	AAO6919	Hexaploid wheat DB
38	9	2.4	9	21	AAO90473	Yeast Sm11 C-termi
39	9	2.4	27	23	AAM51853	P falciparum isop
40	9	2.4	39	22	ABG29276	Novel human diagno
41	9	2.4	44	22	AAO11106	Human polypeptide
42	9	2.4	46	22	AAU31342	Novel human secret
43	9	2.4	78	22	AAO11510	Human polypeptide
44	9	2.4	80	22	AAO09829	Human polypeptide
45	9	2.4	82	22	AAO09044	Human polypeptide

ALIGNMENTS

RESULT 1
AA90470
ID AA90470 standard; Protein: 104 AA.

XX AC AA90470;

XX DT 15-AUG-2000 (first entry)

XX DE Yeast Sm11 (suppressor of mecl lethality) protein.

XX KW Sm11 protein: YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT.

XX OS Saccharomyces cerevisiae.

XX PN WO200017225-A2.

XX PD 30-MAR-2000.
XX PF 24-SEP-1999; 99WO-US22260.
XX PR 24-SEP-1998; 98US-0158858.
XX PA (UYCO) UNIV COLUMBIA NEW YORK.
XX PI Rothstein R, Zhao X;
XX WPI; 2000-283539/24.
XX DR N-PSDB; AAA14263.
XX PT New Sm1 protein and its homologs, useful for treating cancer,
XX PT microbial infection and ataxia telangiectasia and in screening for
XX PT specific modulators
XX PS Claim 2; Fig 1C; 98pp; English.
XX CC This sequence represents the yeast Sm1 (suppressor of mecl lethality)
XX CC protein. This protein is encoded by the SML1 gene, located on chromosome
XX CC XIII, and is an inhibitor of ribonucleotide reductase. The SML1 gene
XX CC (assigned the name YML058w by the Stanford yeast genomic project) was
XX CC initially isolated in a mecl-1 strain by genetic analysis. Sm1 is a
XX CC suppressor of the mecl mutant in yeast which is associated with abnormal
XX CC levels of recombination in both meiosis and mitosis. Sm1 also permits
XX CC cell growth in the absence of the cell cycle checkpoint proteins Mecl
XX CC and Rad53, which, unlike most checkpoint proteins, are essential for cell
XX CC growth. Sm1 binds to the largest subunit of ribonucleotide reductase
XX CC (RNR) which catalyzes the rate-limiting step of deoxynucleotide
XX CC triphosphate (dNTP) synthesis, conversion of NTPs to dNTPs. Binding of
XX CC Sm1 reduces the activity of RNR and thus inhibits the synthesis of
XX CC dNTPs. The SML1 gene, or fragments thereof may be used to isolate the
XX CC human homologue of SML1. The Sm1 protein, and compounds that modulate
XX CC the interaction of Sm1 with ribonucleotide reductase (RNR), may be used
XX CC to alter the rate at which cells divide. These are particularly useful
XX CC for treating cancer, microbial infection and ataxia telangiectasia (AT),
XX CC a condition in which patients are prone to cancer. Sm1 may also be used
XX CC to screen for modulatory agents, to raise specific antibodies, and for
XX CC stimulating the function of the ATM gene (a mammalian Mecl homologue
XX CC which is mutated in AT). Anti-Sm1 antibodies are used as diagnostic and
XX CC analytical immunoassay reagents and to remove Sm1 from serum or to
XX CC titrate Sm1 intracellularly.
XX SQ Sequence 104 AA;
Alignment Scores:
Pred. No.: 6.7e-92 Length: 104
Score: 104.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.73% Indels: 0
DB: 21 Gaps: 0
US-09-814-661a-1 (1-1158) x AAY90470 (1-104)
QY 549 ATGCAAAATCCCAAGACTACTTTTACGCTCAAAATCGTCCCAACACACAGCCCT 608
DB 1 MetGlnAsnSerGlnAspTrpPheTrpAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
QY 609 TCCACATTGCGTACCGTGACCATCGCGGAATTTAGAAGGGTGCGCTTTGCCACCATGCGCT 668
DB 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProProMetAla 40
QY 669 GAGGTTCTCTATGTTGTCTACTCAAAACTCCATGGCAGCTCCGCTTCTGCTCGCTCTCT 728
DB 41 GluValProMetLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
QY 729 TCATTAGAAATGTGGGAAAGGATTTGGAGGAGAGACTCAACTCTATCATCATCATCATG 788
DB 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80

QY 789 AACAAACAACAAATTTGGTGTCTGCGCACTAAATCTATGTTCACACAGGTAAGTCGAG 848
DB 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100
QY 849 GAAATGGACTTC 860
DB 101 GluMetAspPhe 104
RESULT 2
AAB18254
ID AAB18254 standard; Protein: 2206 AA.
XX AC AAB18254;
XX DT 07-NOV-2000 (first entry)
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:111.
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX OS Plasmodium falciparum.
XX PN WO200025728-A2.
XX PD 11-MAY-2000.
XX PF 05-NOV-1999; 99WO-US26796.
XX PR 05-NOV-1998; 98US-0107131.
XX PA (HOFF/) HOFFMAN S.
XX PA (CARU/) CARUCCI D.
XX PA (GARD/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
WPI; 2000-365347/31.
Proteins encoded by chromosome 2 of the human malarial parasite,
Plasmodium falciparum, useful as antimalarial vaccines and in the
diagnosis of P.falciparum infection -
XX Disclosure: Page 263-269; 577pp; English.
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
SQ Sequence 2206 AA;
Alignment Scores:
Pred. No.: 0.0709 Length: 2206
Score: 11.00 Matches: 11

RESULT 5
AAB34410

50 Iletvriletvriletvriletvri 59
pb
pb

```

RESULT 9
AAO12066
ID AAO12066 standard; Protein; 62 AA.
XX
AC AAO12066;
XX
DT 06-NOV-2001. (first entry)
XX
DE Human polypeptide SEQ ID NO 25958.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR N-PSDB; AAI91997.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders.
XX
PS Claim 20; SEQ ID NO 25958; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 62 AA;
XX
Alignment Scores:
Pred. No.: 1.11 Length: 62
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.67% Indels: 0
DB: 22 Gaps: 0
US-09-814-661A-1 (1-1158) x AAO12066 (1-62)
QY 1013 ATATACATATATATATATATATATATATATATATAT 1042
Db 18 IleTyrIleTyrIleTyrIleTyrIleTyr 27
RESULT 10
AAO11381
ID AAO11381 standard; Protein; 67 AA.
XX
AC AAO11381;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 25273.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR N-PSDB; AAI91312.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders.
XX
PS Claim 20; SEQ ID NO 25273; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 67 AA;
XX
Alignment Scores:
Pred. No.: 1.09 Length: 67
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.67% Indels: 0
DB: 22 Gaps: 0
US-09-814-661A-1 (1-1158) x AAO11381 (1-67)
QY 1013 ATATACATATATATATATATATATATATATATATAT 1042
Db 34 IleTyrIleTyrIleTyrIleTyrIleTyr 43
RESULT 11
AAO12482
ID AAO12482 standard; Protein; 67 AA.
XX
AC AAO12482;
XX

```


KW Infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX Homo sapiens.
 OS
 XX
 XX WO200200677-A1.
 PN
 XX
 XX 03-JAN-2002.
 PD
 XX
 XX 07-JUN-2001; 2001WO-US18569.
 XX
 XX
 PF
 XX
 XX
 PR 07-JUN-2000; 2000US-209467P.
 PR
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Birse CE, Rosen CA;
 PI
 XX
 XX WPI: 2002-147878/19.
 DR
 DR N-PSDB; ABQ55360.

XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX

PS Claim 11: SEQ ID NO 3415; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovarian and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 73 AA;

Alignment Scores:

Pred. No.:	1.08	Length:	73
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.67%	Indels:	0
DB:	23	Gaps:	0

US-09-814-661A-1 (1-1158) x ABP42283 (1-73)

OY 1013 ATATACATATATATATATATATATATATATATAT 1042

Db |||||||||||||||||||||||||||||||||||
 22 lleyrlleryrlleryrlleryrlleryr 31

RESULT 14

AAU31237

ID AAU31237 standard; Protein; 75 AA.

XX AAU31237;

AC AAU31237;

DT 18-DEC-2001 (first entry)

XX Novel human secreted protein #1728.

DE Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

XX WO200179449-A2.

PN 25-OCT-2001.

PD 16-APR-2001; 2001WO-US08656.

XX 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-611725/70.

DR Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy -

XX Claim 20: Page 419; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for

CC determining the presence of or predisposition to a disease associated

CC with altered levels of polypeptide. The polypeptides are also useful for

CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent

CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising

CC the nucleic acids encoding the polypeptides and cells genetically

CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and

CC therapy, and can be used as nutritional supplements. They may be used to

CC increase stem cell proliferation; to regulate haematopoiesis; and in

CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC immune suppression and/or stimulation; as anti-inflammatory agents; and

CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid

CC sequences of novel human secreted proteins of the invention.

XX SQ Sequence 75 AA;

Alignment Scores:

Pred. No.:	1.08	Length:	75
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.67%	Indels:	0
DB:	22	Gaps:	0

US-09-814-661A-1 (1-1158) x AAU31237 (1-75)

OY 1013 ATATACATATATATATATATATATATATATATAT 1042

Search completed: July 16, 2003, 12:20:28
Job time : 61 secs

```
Db      64 Iletyrilleyriletyriletyr 73
RESULT 15
AAO0092
ID      AAO00992 standard; Protein: 78 AA.
AC      AAO00992;
DT      06-NOV-2001 (first entry)
DE      Human polypeptide SEQ ID NO 14884.
KW      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW      vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW      tissue growth factor; immunomodulatory; cancer; leukaemia;
KW      nervous system disorders; arthritis; inflammation.
OS      Homo sapiens.
PN      WO200164835-A2.
PD      07-SEP-2001.
PF      26-FEB-2001; 2001WO-US04927.
PR      28-FEB-2000; 2000US-0515126.
PR      18-MAY-2000; 2000US-0577409.
PA      (HYSE-) HYSEQ INC.
PI      Tang YT, Liu C, Drmanac RT;
DR      WPI: 2001-514838/56.
DR      N-PSDB; AAI80923.
PT      Isolated nucleic acids and polypeptides, useful for preventing
PT      diagnosing and treating e.g. leukaemia, inflammation and immune
PT      disorders -
XX      Claim 20; SEQ ID NO 14884; 1399pp + Sequence Listing; English.
XX      The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX      the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
XX      cytokine, cell proliferation or cell differentiation or which may induce
XX      production of other cytokines in other cell populations. The
XX      polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX      peptide therapy. The polypeptides have various cytokine-like activities,
XX      e.g. stem cell growth factor activity, haematopoiesis regulating
XX      activity, tissue growth factor activity, immunomodulatory activity and
XX      activin/inhibin activity and may be useful in the diagnosis and/or
XX      treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX      inflammation.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX      Sequence 78 AA;
Alignment Scores:
Pred. No.:      1.07      Length:      78
Score:          10.00      Matches:      10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      2.67%      Indels:      0
DB:              22      Gaps:      0
US-09-814-661A-1 (1-1158) x AAO00992 (1-78)
QY      1013 ATATACATATATATATATATATAT 1042
Db      43 Iletyrilleyriletyriletyr 52
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 16, 2003, 12:18:40 ; Search time 17 Seconds
(without alignments)
4008.440 Million cell updates/sec

Title: US-09-814-661A-1

Perfect score: 375

Sequence: 1 aatgagcaacggtgcaaca.....cctcttctgataatgcttt 1158

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spo1/US09814661/runat_16072003_115332_15244/app_query.fasta_1.1351
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09814661@cgn_1.1.28 @runat_16072003_115332_15244 -NCPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/pckfiles1.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description
1	9	2.4	2391	2	US-08-446-855A-2
2	9	2.4	2391	4	US-09-150-741-2
3	9	2.4	2710	2	US-08-568-459A-12
4	9	2.4	2710	2	US-08-487-826B-12
5	9	2.4	2710	4	US-09-210-288-12
6	8	2.2	10	1	US-08-468-674B-86
7	8	2.2	10	1	US-08-468-674B-88
8	8	2.2	10	1	US-08-780-571-86
9	8	2.2	10	1	US-08-780-571-88
10	8	2.2	10	4	US-09-012-669F-45
11	8	2.2	13	4	US-08-932-082-6
12	8	2.2	54	1	US-08-463-660-4

c 13	8	2.2	54	1	US-08-678-280-4	Sequence 4, Appli
c 14	8	2.2	65	1	US-08-468-674B-71	Sequence 71, Appl
c 15	8	2.2	65	1	US-08-780-571-71	Sequence 71, Appl
c 16	8	2.2	88	1	US-08-468-674B-75	Sequence 75, Appl
c 17	8	2.2	88	1	US-08-780-571-75	Sequence 75, Appl
c 18	8	2.2	146	1	US-08-400-250-48	Sequence 48, Appl
c 19	8	2.2	146	3	US-08-975-365-48	Sequence 48, Appl
c 20	8	2.1	523	3	US-08-606-505B-67	Sequence 67, Appl
c 21	8	2.1	523	4	US-09-616-990-67	Sequence 67, Appl
c 22	8	2.1	651	1	US-08-431-080-24	Sequence 24, Appl
c 23	8	2.1	651	2	US-08-938-534-24	Sequence 24, Appl
c 24	8	2.1	651	4	US-09-345-294-24	Sequence 24, Appl
c 25	8	2.1	738	3	US-08-864-038A-3	Sequence 3, Appli
c 26	8	2.1	909	4	US-09-425-383-2	Sequence 3, Appli
c 27	8	2.1	986	2	US-08-673-789-3	Sequence 3, Appli
c 28	8	2.1	1018	1	US-08-408-093-6	Sequence 6, Appli
c 29	8	2.1	1018	1	US-08-408-420A-6	Sequence 6, Appli
c 30	8	2.1	1018	1	US-08-714-901-6	Sequence 6, Appli
c 31	8	2.1	1018	3	US-08-040-741-6	Sequence 6, Appli
c 32	8	2.2	2391	2	US-08-446-855A-2	Sequence 2, Appli
c 33	8	2.2	2391	4	US-09-150-741-2	Sequence 2, Appli
c 34	8	2.2	2710	2	US-08-568-459A-12	Sequence 12, Appl
c 35	8	2.2	2710	2	US-08-487-826B-12	Sequence 12, Appl
c 36	8	2.2	2710	4	US-09-210-288-12	Sequence 12, Appl
c 37	7	1.9	9	1	US-08-468-674B-85	Sequence 85, Appl
c 38	7	1.9	9	1	US-08-468-674B-87	Sequence 87, Appl
c 39	7	1.9	9	1	US-08-780-571-85	Sequence 85, Appl
c 40	7	1.9	9	1	US-08-780-571-87	Sequence 87, Appl
c 41	7	1.9	10	5	PCT-US95-16415-21	Sequence 21, Appl
c 42	7	1.9	15	4	US-08-218-369-1	Sequence 1, Appli
c 43	7	1.9	15	5	PCT-US95-03742-1	Sequence 1, Appli
c 44	7	1.9	16	1	US-08-346-849-26	Sequence 26, Appl
c 45	7	1.9	16	2	US-08-293-284A-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1

US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No, 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29.009
; REFERENCE/DOCKET NUMBER: 47-80
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 2391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-855A-2

Alignment Scores:
Pred. No.: 1.12 Length: 2391
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.40% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-446-855A-2 (1-2391)
Qy 1016 TACATATATATATATATATATATAT 1042
Db 10 TyrIleTyrIleTyrIleTyrIleTyr 18

RESULT 2
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; FILING DATE: 1992-12-16
; PUBLICATION DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446.855
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Alignment Scores:
Pred. No.: 1.12 Length: 2391
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.40% Indels: 0
DB: 4 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-150-741-2 (1-2391)
Qy 1016 TACATATATATATATATATATATAT 1042
Db 10 TyrIleTyrIleTyrIleTyrIleTyr 18

RESULT 3
US-08-568-459A-12
; Sequence 12, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US

; APPLICANT: Wellem's, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568.459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel'sen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-568-459A-12

Alignment Scores:
Pred. No.: 1.1 Length: 2710
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.40% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-568-459A-12 (1-2710)
Qy 1021 ATATATATATATATATATATATGTC 1047
Db 2653 IleTyrIleTyrIleTyrIleCysLeu 2661

RESULT 4
US-08-487-826B-12
; Sequence 12, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem's, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
```

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Alignment Scores:
Pred. No.: 1.1 Length: 2710
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.40% Indels: 0
DB: 2 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-487-826B-12 (1-2710)

Qy 1021 ATATATATATATATATATATGTCCTC 1047
|||||
Db 2653 IleTyrIleTyrIleTyrIleCysLeu 2661

RESULT 5
US-09-210-288-12
Sequence 12, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellem, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORGANISM: Plasmodium falciparum
US-09-210-288-12

Alignment Scores:
Pred. No.: 1.1 Length: 2710
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.40% Indels: 0
DB: 4 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-210-288-12 (1-2710)

Qy 1021 ATATATATATATATATATATGTCCTC 1047
|||||
Db 2653 IleTyrIleTyrIleTyrIleCysLeu 2661

RESULT 6

US-08-468-674B-86

Sequence 86, Application US/08468674B

Patent No. 5639642

GENERAL INFORMATION:

APPLICANT: Kjeldsen, Thomas B

APPLICANT: Vad, Knud

TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES

NUMBER OF SEQUENCES: 89

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5639642 No. 5639642disk of No. 5639642th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Tape

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,674B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/282,852

FILING DATE: 29-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4085,220-US

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 86:

SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids

TYPE: amino acid

STRANDEDNESS: single

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-468-674B-86

Alignment Scores:
Pred. No.: 22.7 Length: 10
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-468-674B-86 (1-10)
Qy 730 GAAGAGCGGAGCGAGCGGAG 707
Db 1 GluGluAlaGluAlaGluAlaGlu 8

RESULT 7
US-08-468-674B-88
; Sequence 88, Application US/08468674B
; Patent No. 5639642
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,674B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 33,728
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/282,852
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4085.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-468-674B-88

Alignment Scores:
Pred. No.: 22.7 Length: 10
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-468-674B-88 (1-10)
Qy 730 GAAGAGCGGAGCGAGCGGAG 707
Db 1 GluGluAlaGluAlaGluAlaGlu 8

RESULT 8
US-08-780-571-86
; Sequence 86, Application US/08780571
; Patent No. 5795746
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5795746o No. 5795746disk of No. 5795746th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/780,571
; FILING DATE: 08-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/468,674
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/282,852
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4085.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-780-571-86

Alignment Scores:
Pred. No.: 22.7 Length: 10
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-780-571-86 (1-10)
Qy 730 GAAGAGCGGAGCGAGCGGAG 707
Db 1 GluGluAlaGluAlaGluAlaGlu 8

RESULT 9
US-08-780-571-88
; Sequence 88, Application US/08780571
; Patent No. 5795746
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
```


ADDRESSER: No. 57957460 No. 5795746disk of No. 5795746th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor

CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Tape

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/780-571

FILING DATE: 08-JAN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/468,674

FILING DATE: 06-JUN-1995

APPLICATION NUMBER: US 08/282,852

FILING DATE: 29-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4085.220-US

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 88:

SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-780-571-88

Alignment Scores:
Pred. No.: 22.7 Length: 10
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-780-571-88 (1-10)

QY 730 GAAGAAGCGGAGCGAGCGGAG 707

Db 1 GluGluAlaGluAlaGluAlaGlu 8

RESULT 10

US-09-012-669F-45

Sequence 45, Application US/09012669F

Patent No. 6214547

GENERAL INFORMATION:

APPLICANT: Kjelidsen, Thomas B.

APPLICANT: Havelund, Svend

APPLICANT: Pettersson, Annette F.

APPLICANT: Balschmidt, Per

TITLE OF INVENTION: Synthetic Leader Peptide Sequences

FILE REFERENCE: 4954.200-US

CURRENT APPLICATION NUMBER: US/09/012,669F

PRIOR FILING DATE: 1998-01-23

PRIOR APPLICATION NUMBER: 0092/97

PRIOR FILING DATE: 1997-01-24

PRIOR APPLICATION NUMBER: PCT/DK98/00026

PRIOR FILING DATE: 1998-01-22

NUMBER OF SEQ ID NOS: 57

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 45

LENGTH: 10

TYPE: PRT

ORGANISM: S. cerevisiae

US-09-012-669F-45

Alignment Scores:

Pred. No.: 22.7 Length: 10

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.16% Indels: 0

DB: 4 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-012-669F-45 (1-10)

QY 730 GAAGAAGCGGAGCGAGCGGAG 707

Db 1 GluGluAlaGluAlaGluAlaGlu 8

RESULT 11

US-08-932-082-6

Sequence 6, Application US/08932082

Patent No. 6251856

GENERAL INFORMATION:

APPLICANT: Markussen, Jan

APPLICANT: Jonassen, Ib

APPLICANT: Havelund, Svend

APPLICANT: Brandt, Jakob

APPLICANT: Kurtzhals, Peter

APPLICANT: Hansen, Hertz Per

APPLICANT: Kaarsholm, Niels Christian

TITLE OF INVENTION: INSULIN DERIVATIVES

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 62518560 No. 6251856disk of No. 6251856th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/932,082

FILING DATE: 12-AUG-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4341.204-US

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 13 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-932-082-6

Alignment Scores:

Pred. No.: 21.9 Length: 13

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.16% Indels: 0

DB: 4 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-932-082-6 (1-13)

QY 730 GAAGAAGCGGAGCGAGCGGAG 707

Db 1 GluGluAlaGluAlaGluAlaGlu 8

Db 1 GluGluAlaGluAlaGluAlaGlu 8

RESULT 12

US-08-463-660-4

; Sequence 4, Application US/08463660

; Patent No. 5759776

; GENERAL INFORMATION:

; APPLICANT: SMITH, HELENE S.

; APPLICANT: CHUN, LING-CHUN

; TITLE OF INVENTION: TARGETS FOR BREAST CANCER DIAGNOSIS AND TREATMENT

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/463,660

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: CIOTTI, THOMAS E.

; REGISTRATION NUMBER: 21,013

; REFERENCE/DOCKET NUMBER: 28888-20001.00

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-463-660-4

Alignment Scores:

Pred. No.: 18 Length: 54

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.16% Indels: 0

DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-463-660-4 (1-54)

OY 942 AAAGAAGGGTATCTAAGAGAGAA 919

|||||

Db 22 LysGluGlyTyrLeuArgGluGlu 29

RESULT 13

US-08-678-280-4

; Sequence 4, Application US/08678280

; Patent No. 5776683

; GENERAL INFORMATION:

; APPLICANT: SMITH, HELENE S.

; APPLICANT: CHUN, LING-CHUN

; TITLE OF INVENTION: TARGETS FOR BREAST CANCER DIAGNOSIS AND

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/463,660

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: CIOTTI, THOMAS E.

; REGISTRATION NUMBER: 21,013

; REFERENCE/DOCKET NUMBER: 28888-20001.00

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-463-660-4

Alignment Scores:

Pred. No.: 18 Length: 54

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.16% Indels: 0

DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-463-660-4 (1-54)

OY 942 AAAGAAGGGTATCTAAGAGAGAA 919

|||||

Db 22 LysGluGlyTyrLeuArgGluGlu 29

RESULT 14

US-08-468-674B-71

; Sequence 71, Application US/08468674B

; Patent No. 5639642

; GENERAL INFORMATION:

; APPLICANT: Kjeldsen, Thomas B

; APPLICANT: Vad, Knud

; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES

; NUMBER OF SEQUENCES: 89

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Tape

; COMPUTER: IBM PC compatible.

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/468,674B

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/282,852

; FILING DATE: 29-JUL-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4085.220-US

; TELECOMMUNICATION INFORMATION:

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/678,280

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Schiff, J. Michael

; REGISTRATION NUMBER: 40,253

; REFERENCE/DOCKET NUMBER: 28888-20001.20

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; TELEFAX: 706141

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-678-280-4

Alignment Scores:

Pred. No.: 18 Length: 54

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.16% Indels: 0

DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-678-280-4 (1-54)

OY 942 AAAGAAGGGTATCTAAGAGAGAA 919

|||||

Db 22 LysGluGlyTyrLeuArgGluGlu 29

RESULT 14

US-08-468-674B-71

; Sequence 71, Application US/08468674B

; Patent No. 5639642

; GENERAL INFORMATION:

; APPLICANT: Kjeldsen, Thomas B

; APPLICANT: Vad, Knud

; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES

; NUMBER OF SEQUENCES: 89

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Tape

; COMPUTER: IBM PC compatible.

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/468,674B

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/282,852

; FILING DATE: 29-JUL-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4085.220-US

; TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-674B-71

Alignment Scores:
Pred. No.: 17.5 Length: 65
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-468-674B-71 (1-65)

QY 730 GAAGAAGCGGAGCGAGCGGAG 707
|||||
Db 3 GluGluAlaGluAlaGluAlaGlu 10

RESULT 15

US-08-780-571-71
; Sequence 71, Application US/08780571
; Patent No. 5795746
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 57957460 No. 5795746disk of No. 5795746th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/780,571
; FILING DATE: 08-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/468,674
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/282,852
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4085.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-780-571-71

Alignment Scores:
Pred. No.: 17.5 Length: 65
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.16% Indels: 0
DB: 1 Gaps: 0

US-09-814-661A-1 (1-1158) x US-08-780-571-71 (1-65)

QY 730 GAAGAAGCGGAGCGAGCGGAG 707
|||||
Db 3 GluGluAlaGluAlaGluAlaGlu 10

Search completed: July 16, 2003, 12:26:16
Job time : 31 secs

THIS PAGE BLANK (USPTO)


```

US-09-814-661A-1 (1-1158) x US-09-814-661A-2 (1-104)

QY 549 ATGCAAAATCCCAAGACTACTTTTAGCTCAAAATCGCTGCCAACAAACAAGCCCT 608
Db 1 MetGlnAsnSerGlnAspTyrPheTyrAlaGlnAsnArgCysGlnGlnGlnAlaPro 20
QY 609 TCCACATTCGCTACCGTGACCATGCCGAATTTAGAAGGTGCTTTGCCACCTATGGCT 668
Db 21 SerThrLeuArgThrValThrMetAlaGluPheArgValProLeuProPrometAla 40
QY 669 GAGGTTCCTATGTTGCTGTACTCAAAACTCCCATGGCAGCTCCGCTTCTGCCTCCGCTTCT 728
Db 41 GluValPrometLeuSerThrGlnAsnSerMetGlySerSerAlaSerAlaSerAlaSer 60
QY 729 TCATTAGAAATGTGGGAAAGGATTGTGAGAGAGACTCAACTCTCTATCGATCATGACATG 788
Db 61 SerLeuGluMetTrpGluLysAspLeuGluGluArgLeuAsnSerIleAspHisAspMet 80
QY 789 AACAAACAACAAATTTGGTTCGTGGCGAACTAAAATCTATGTTCAACAGGGTAAGGTCCGAG 848
Db 81 AsnAsnAsnLysPheGlySerGlyGluLeuLysSerMetPheAsnGlnGlyLysValGlu 100

QY 849 GAAATGGACTTC 860
Db 101 GluMetAspPhe 104

RESULT 2
US-10-275-360-6
; Sequence 6, Application US/10275360
; Publication No. US20030115634A1
; GENERAL INFORMATION:
; APPLICANT: JOWAA, Hassan
; TITLE OF INVENTION: GENES OF THE 1-DESOXY-D-XYLULOSE BIOSYNTHESIS PATH
; FILE REFERENCE: JOWAA-12 (PCT)
; CURRENT APPLICATION NUMBER: US/10/275,360
; CURRENT FILING DATE: 2002-11-05
; PRIOR APPLICATION NUMBER: PCT/EP01/04537
; PRIOR FILING DATE: 2001-04-21
; PRIOR APPLICATION NUMBER: DE10021688.9
; PRIOR FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-275-360-6

Alignment Scores:
Pred. No.: 0.985 Length: 19
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.67% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-10-275-360-6 (1-19)

QY 1013 ATATACATATATATATATATATATATATATATATATATATATATATATATATAT 1042
Db 9 IleTyrIleTyrIleTyrIleTyrIleTyr 18

RESULT 3
US-10-275-360-14
; Sequence 14, Application US/10275360
; Publication No. US20030115634A1
; GENERAL INFORMATION:
; APPLICANT: JOWAA, Hassan
; TITLE OF INVENTION: GENES OF THE 1-DESOXY-D-XYLULOSE BIOSYNTHESIS PATH
; FILE REFERENCE: JOWAA-12 (PCT)
; CURRENT APPLICATION NUMBER: US/10/275,360
; CURRENT FILING DATE: 2002-11-05
; PRIOR APPLICATION NUMBER: PCT/EP01/04537

```

APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,291A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 285
LENGTH: 463
TYPE: PRT
ORGANISM: Homo Sapien
US-09-905-291A-285
Alignment Scores: 0.635 Length: 463
Pred. No.: 10.00 Matches: 10
Score:

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0
US-09-814-661A-1 (1-1158) x US-09-905-291A-285 (1-463)
QY 730 GAAGAAGCGAGGAGCGAGAGCGAGCTGCC 701
|||||
Db 355 GUGluAlaGluAlaGluAlaGluLeuPro 364
RESULT 6
US-09-902-853-285
; Sequence 285, Application US/099028553
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 285
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-905-291A-285

; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 285
 ; LENGTH: 463
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-902-853-285

Alignment Scores:
 Pred. No.: 0.635 Length: 463
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.70% Indels: 0
 DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-902-853-285 (1-463)

QY 730 GAAGAGCGGAGCGAGCGAGCGAGCTGCC 701
 Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 7

US-09-907-824-285
 ; Sequence 285, Application US/09907824
 ; Publication No. US20020197671A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kljavin, Ivar J.
 ; APPLICANT: Mather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William, I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; FILE REFERENCE: 10466-14
 ; CURRENT APPLICATION NUMBER: US/09/907,824
 ; CURRENT FILING DATE: 2001-07-17
 ; PRIOR APPLICATION NUMBER: 09/665,350
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414
 ; PRIOR FILING DATE: 2000-02-22
 ; PRIOR APPLICATION NUMBER: US 60/143,048
 ; PRIOR FILING DATE: 1999-07-07
 ; PRIOR APPLICATION NUMBER: US 60/145,698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1999-07-28

; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 285
 ; LENGTH: 463
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-907-824-285

Alignment Scores:
 Pred. No.: 0.635 Length: 463
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.70% Indels: 0
 DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-907-824-285 (1-463)

QY 730 GAAGAGCGGAGCGAGCGAGCGAGCTGCC 701
 Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 8

US-09-907-841-285
 ; Sequence 285, Application US/09907841
 ; Publication No. US20020198366A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kljavin, Ivar J.
 ; APPLICANT: Mather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,841
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 285
LENGTH: 463
TYPE: PRT
ORGANISM: Homo Sapien
US-09-907-841-285

Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-907-841-285 (1-463)

QY 730 GAAGAAGCGGAGCGAGCGAGCGAGCGCC 701
Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 9
US-09-904-011-285
Sequence 285, Application US/09904011
Publication No. US20030003530A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Oiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.

APPLICANT: Mather, Jennie P.
APPLICANT: Paoni, James
APPLICANT: Pao, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/904,011
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30959
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 285
LENGTH: 463
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-011-285

Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-904-011-285 (1-463)

QY 730 GAAGAAGCGGAGCGAGCGAGCGAGCGCC 701
Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 10
US-10-028-072-360
Sequence 360, Application US/10028072

Publication No. US20030004311A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Naureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OF INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028,072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063550
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149

; GENERAL INFORMATION:

US-09-906-742-285

US-09-906-742-285

```
Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-906-742-285 (1-463)
QY 730 GAAGAAGCGGAGCAGACGAGCTGCC 701
Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 12
US-10-121-049-360
; Sequence 360, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 360
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-360

Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-10-121-049-360 (1-463)
QY 730 GAAGAAGCGGAGCAGACGAGCTGCC 701
Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 13
US-10-123-904-360
; Sequence 360, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
```

```
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 360
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-904-360

Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-10-123-904-360 (1-463)
QY 730 GAAGAAGCGGAGCAGACGAGCTGCC 701
Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 14
US-10-140-470-360
; Sequence 360, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 360
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-360
```

Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-10-140-470-360 (1-463)

QY 730 GAAGAAGCGGAGCGAGCGAGCGGCTGCC 701

Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

RESULT 15

US-09-906-838-285
Sequence 285, Application US/09906838
Publication No. US20030027143A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/906,838

PRIOR FILING DATE: 2001-07-16

PRIOR APPLICATION NUMBER: 09/665,350

PRIOR FILING DATE: 2000-09-18

PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: US 60/143,048

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: US 60/145,698

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: US 60/146,222

PRIOR FILING DATE: 1999-07-28

PRIOR APPLICATION NUMBER: PCT/US99/20594

PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: PCT/US99/21090

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/21547

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/23089

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: PCT/US99/28214

PRIOR FILING DATE: 1999-11-29

PRIOR APPLICATION NUMBER: PCT/US99/28313

PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 285
LENGTH: 463
TYPE: PRT
ORGANISM: Homo Saplen
US-09-906-838-285
Alignment Scores:
Pred. No.: 0.635 Length: 463
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.70% Indels: 0
DB: 9 Gaps: 0

US-09-814-661A-1 (1-1158) x US-09-906-838-285 (1-463)

QY 730 GAAGAAGCGGAGCGAGCGAGCGGCTGCC 701

Db 355 GluGluAlaGluAlaGluAlaGluLeuPro 364

Search completed: July 16, 2003, 12:41:17

Job time : 54.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:44:31 ; Search time 50 Seconds
(without alignments)
247.020 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 537

Sequence: 1 MQNSQDYFYAQNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	537	100.0	104	10	US-09-814-661A-2
2	68.5	12.8	785	10	US-09-801-368-348
3	68.5	12.8	1068	10	US-09-801-368-242
4	68	12.7	486	10	US-09-801-368-398
5	68	12.7	1019	9	US-10-270-333-93
6	68	12.7	2843	8	US-08-681-219-32
7	68	12.7	2843	9	US-09-987-482-1
8	67.5	12.6	209	12	US-10-124-557-94
9	67.5	12.6	220	12	US-10-124-557-96
10	67.5	12.6	260	9	US-10-156-761-13855
11	67.5	12.6	1140	12	US-10-124-557-104
12	67.5	12.6	1363	12	US-10-124-557-52
13	67.5	12.6	1404	12	US-10-124-557-2
14	67.5	12.6	1404	12	US-10-124-557-62
15	67	12.5	348	10	US-09-835-996A-17
16	67	12.5	348	10	US-09-835-996A-43
17	66.5	12.4	1115	9	US-10-205-823-160
18	66.5	12.4	1115	10	US-09-781-558-2
19	65.5	12.2	185	10	US-09-881-752A-40

20	65.5	12.2	391	12	US-10-062-254-212	Sequence 212, App
21	65	12.1	480	9	US-09-893-519A-9	Sequence 9, Appli
22	65	12.1	721	10	US-09-945-258-12	Sequence 12, Appl
23	64.5	12.0	556	9	US-10-219-220-259	Sequence 259, App
24	64.5	12.0	714	10	US-09-529-063-16	Sequence 16, Appl
25	64	11.9	356	12	US-10-062-254-262	Sequence 262, App
26	64	11.9	675	9	US-09-928-530-2	Sequence 2, Appli
27	64	11.9	675	9	US-10-162-012-27	Sequence 27, Appli
28	64	11.9	675	9	US-09-733-630-2	Sequence 2, Appli
29	64	11.9	1569	9	US-10-108-605-303	Sequence 303, App
30	63	11.7	221	10	US-09-801-368-432	Sequence 432, App
31	62.5	11.6	208	12	US-10-124-557-132	Sequence 132, App
32	62.5	11.6	419	10	US-09-814-777A-36	Sequence 36, Appl
33	62.5	11.6	434	10	US-09-866-582-14	Sequence 14, Appl
34	62.5	11.6	710	10	US-09-815-242-10895	Sequence 10895, A
35	62.5	11.6	790	10	US-09-801-368-228	Sequence 228, App
36	62	11.5	159	10	US-09-864-761-35248	Sequence 35248, A
37	62	11.5	514	10	US-09-922-138-5	Sequence 5, Appli
38	62	11.5	514	10	US-09-841-683-2	Sequence 2, Appli
39	62	11.5	1150	9	US-10-108-605-301	Sequence 301, App
40	62	11.5	1191	9	US-10-087-464-47	Sequence 47, Appl
41	61.5	11.5	211	9	US-09-738-626-6733	Sequence 6733, Ap
42	61.5	11.5	459	9	US-09-468-147-206	Sequence 206, App
43	61.5	11.5	459	9	US-09-468-147-207	Sequence 207, App
44	61.5	11.5	1237	9	US-10-108-605-211	Sequence 211, App
45	61.5	11.5	1394	9	US-10-108-605-213	Sequence 213, App

ALIGNMENTS

RESULT 1

US-09-814-661A-2
; Sequence 2, Application US/09814661A
; Patent No. US20020151016A1

; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney

; APPLICANT: Zhao, Xiaolan

; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE

; FILE REFERENCE: 0575/56615-A-PCT-US

; CURRENT APPLICATION NUMBER: US/09/814,661A

; CURRENT FILING DATE: 2001-03-22

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 104

; TYPE: PRT

; ORGANISM: S. Cerevisiae

; US-09-814-661A-2

Query Match 100.0%; Score 537; DB 10; Length 104;
Best Local Similarity 100.0%; Pred. No. 3.5e-54;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQNSQDYFYAQNRCQQQAPSTLRTVTMAEVRVPLPPMAEVPMLSTQNSMGSSASAS 60

Db 1 MQNSQDYFYAQNRCQQQAPSTLRTVTMAEVRVPLPPMAEVPMLSTQNSMGSSASAS 60

Qy 61 SLEWKEKDLERLNSIDHDMMNNKFGSGELKSMFNOGKVEEMDF 104

Db 61 SLEWKEKDLERLNSIDHDMMNNKFGSGELKSMFNOGKVEEMDF 104

RESULT 2

US-09-801-368-348

; Sequence 348, Application US/09801368

; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

```

; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 348
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-348

Query Match      12.8%; Score 68.5; DB 10; Length 785;
Best Local Similarity 24.7%; Pred. No. 34;
Matches 24; Conservative 19; Mismatches 31; Indels 23; Gaps 5;

QY 5 QDYFAQNR--CQOQOASTLRTVTMAEFRRVPLPPMAEVPMLSTQNSMGSSASASL 62
DB 355 QETYLTAHRXGSDQYDSMAKTMNSFTTIRHP-----MPLIATTNATGNTSOTSASI 410
QY 63 -----ENWEXD-----LEERLNSI-----DHDNHN 82
DB 411 IRRVTTMWEDEKTLCTCYQVEANGISVVRANDNMVN 447

RESULT 3
US-09-801-368-242
; Sequence 242, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 242
; LENGTH: 1068
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-242

Query Match      12.8%; Score 68.5; DB 10; Length 1068;

```

```

Best Local Similarity 17.2%; Pred. No. 51;
Matches 22; Conservative 27; Mismatches 28; Indels 51; Gaps 4;

QY 17 QOAPSTLRTVTMAEFRRVP-----LPP 38
DB 938 RSAPSPVRNNSLPEPAQLPSFRSLSVSDMINPDYAQTNGQNTQVQSNKPINAOQQIPT 997
QY 39 MAEVPMLSTQNSMGSSASASASLSEWKEKLEERLNSIDHDNHN-----KEGSGELKSMF 94
DB 998 SVQVPMNT--NEINNNNNNNNN-----KNNINNNNNNSNFSATSPNLGTLDEFV 1048
QY 95 NOGKVEEM 102
DB 1049 NNGDLEDL 1056

RESULT 4
US-09-801-368-398
; Sequence 398, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 398
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-398

Query Match      12.7%; Score 68; DB 10; Length 486;
Best Local Similarity 25.7%; Pred. No. 20;
Matches 26; Conservative 14; Mismatches 37; Indels 24; Gaps 3;

QY 18 QAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSMGSSASASLSEW-----EKDL 69
DB 85 KSPSAKAFTDKQRKN-----EVPNISVSNVFPQSSSTTSSTTSGCCKWSEKVE 137
QY 70 ERLNSIDHDNHN-----KEGSGELKSMFNOGKVEE 101
DB 138 EAFLEALRLKNGTTKIRNANFRNELISLYIKHKTNE 178

RESULT 5
US-10-270-333-93
; Sequence 93, Application US/10270333
; Publication No. US20030092124A1
; GENERAL INFORMATION:
; APPLICANT: Cravchik, Anibal
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF AS INSECTICIDAL TARGETS
; FILE REFERENCE: CL000733CON

```



```
; CURRENT APPLICATION NUMBER: US/10/270,333
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/168,677
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/175,691
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/191,638
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Drosophila
US-10-270-333-93

Query Match          12.7%; Score 68; DB 9; Length 1019;
Best Local Similarity 29.0%; Pred. No. 54;
Matches 27; Conservative 12; Mismatches 48; Indels 6; Gaps 3;

QY 14 CQQQAPSTLRVTMAEFRVPLPPMAEVPMLSTON-SMGSSASASASLEWWEKDLER 72
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 784 CPHRRASSPMETSSKDAIILSPMPPLSPKVKTSPTQGTQSAAGVTPPVPLRIER 843
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 73 LNSIDHMNNKF--GSGELKSMFNOGKVEEMD 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 844 ETS---DSNTNKMPSSTSGTSGVGGGEQD 873
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-681-219-32
; Sequence 32, Application US/08681219
; Patent No. US20020058607A1
; GENERAL INFORMATION:
; APPLICANT: Takaaki Sato and Junn Yanagisawa
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGE
; TITLE OF INVENTION: (PDZ/DHR) DOMAIN AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk.
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,219
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/48962/JPHW/JKM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-681-219-32

Query Match          12.7%; Score 68; DB 8; Length 2843;
Best Local Similarity 27.8%; Pred. No. 2.2e+02;

Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRVTMAEFRVPLPPMAEVPMLSTONSGSSA---SASASLEWWEKDLERLN 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2534 ESPSRLPINRSGTWKREHSHSSLPVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 75 STDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2594 SI-----SGTKQSKENQVSAKGTWRKIKENE 2620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-09-987-482-1
; Sequence 1, Application US/09987482
; Publication No. US20020184656A1
; GENERAL INFORMATION:
; APPLICANT: BHANDARI, POONAM
; APPLICANT: SHASHIDHARA, L.S.
; TITLE OF INVENTION: IN VIVO ASSAY SYSTEM FOR SCREENING AND VALIDATION OF
; TITLE OF INVENTION: DRUGS AND OTHER SUBSTANCES
; FILE REFERENCE: 056859-0134
; CURRENT APPLICATION NUMBER: US/09/987,482
; CURRENT FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2843
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-987-482-1

Query Match          12.7%; Score 68; DB 9; Length 2843;
Best Local Similarity 27.8%; Pred. No. 2.2e+02;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;

QY 18 QAPSTLRVTMAEFRVPLPPMAEVPMLSTONSGSSA---SASASLEWWEKDLERLN 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2534 ESPSRLPINRSGTWKREHSHSSLPVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 75 STDHDMNNKFGSGELKSMFNO-----GKVEEMDF 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2594 SI-----SGTKQSKENQVSAKGTWRKIKENE 2620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-10-124-557-94
; Sequence 94, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; APPLICANT: Clark, Stephen C.
; APPLICANT: Jacobs, Kenneth
; APPLICANT: Hewick, Rodney M.
; APPLICANT: Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-10-124-557-94
Query Match 12.6% Score 67.5; DB 12; Length 209;
Best Local Similarity 23.6%; Pred. No. 7.4;
Matches 21; Conservative 17; Mismatches 24; Indels 27; Gaps 3;

QY 19 APSTLRTVTMAEFRRVPLPP-----MAEVPMLSTQNSMGSSASASASLEMM 65
DB 125 ASQTIKSTT-----KRSPKPPNKKTKKVEIESEITEHSVSENQESSSSSSSSSTIW 180

QY 66 E-----KDLEERLNSIDHDMNNK 84
DB 181 KIKSKNSAANRELQKLVKDKNNRTK 209

RESULT 9
US-10-124-557-96
; Sequence 96, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901

```

```

; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 220 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 96:
US-10-124-557-96
Query Match 12.6% Score 67.5; DB 12; Length 220;
Best Local Similarity 23.6%; Pred. No. 7.9;
Matches 21; Conservative 17; Mismatches 24; Indels 27; Gaps 3;

QY 19 APSTLRTVTMAEFRRVPLPP-----MAEVPMLSTQNSMGSSASASASLEMM 65
DB 125 ASQTIKSTT-----KRSPKPPNKKTKKVEIESEITEHSVSENQESSSSSSSSSTIW 180

QY 66 E-----KDLEERLNSIDHDMNNK 84
DB 181 KIKSKNSAANRELQKLVKDKNNRTK 209

RESULT 10
US-10-156-761-13855
; Sequence 13855, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13855
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-13855

Query Match 12.6% Score 67.5; DB 9; Length 260;
Best Local Similarity 31.7%; Pred. No. 9.9;
Matches 19; Conservative 10; Mismatches 22; Indels 9; Gaps 2;

QY 41 EYVPMSTQNSMGSSASA---SASSLEMMWEKDLLEERLNSIDHDMN-----NNKFGSGELK 91
DB 95 EYVGVSVSNFLRTSATAIRGEAPKLTWVKEYKEAMERLNMDFAPFAERNVNEGFSGGEKK 154

RESULT 11
US-10-124-557-104
; Sequence 104, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.

```

```

; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
;
; US-10-124-557-104
;
; Query Match 12.6%; Score 67.5; DB 12; Length 1140;
; Best Local Similarity 23.6%; Pred. No. 72; Mismatches 24; Indels 27; Gaps 3;
; Matches 21; Conservative 17;
;
; QY 19 APSTLRTVTMAEFRVRVPLPP-----MAEVPMLSTQNSMGSSASASASLEMW 65
; | : : : : | : : : : | : : : : | : : : : | : : : : |
; Db 125 ASQTIKSTT-----KRSPPKPKKTKKVVESIEITEHSVSENQSSSSSSSSSTIW 180
;
; QY 66 E-----KDLERLNSIDHDMNNK 84
; : : : : : | : : : : | : : : : | : : : : | : : : : |
; Db 181 KIKSKNSAANRELQKLVKDKNKNRTK 209
;
; RESULT 12
; US-10-124-557-52
; Sequence 52, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.

```

```

; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
;
; US-10-124-557-52
;
; Query Match 12.6%; Score 67.5; DB 12; Length 1363;
; Best Local Similarity 23.6%; Pred. No. 92; Mismatches 24; Indels 27; Gaps 3;
; Matches 21; Conservative 17;
;
; QY 19 APSTLRTVTMAEFRVRVPLPP-----MAEVPMLSTQNSMGSSASASASLEMW 65
; | : : : : | : : : : | : : : : | : : : : | : : : : |
; Db 84 ASQTIKSTT-----KRSPPKPKKTKKVVESIEITEHSVSENQSSSSSSSSSTIW 139
;
; QY 66 E-----KDLERLNSIDHDMNNK 84
; : : : : : | : : : : | : : : : | : : : : | : : : : |
; Db 140 KIKSKNSAANRELQKLVKDKNKNRTK 168
;
; RESULT 13
; US-10-124-557-2
; Sequence 2, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
MOLECULE TYPE: protein
LENGTH: 1404 amino acids
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE CHARACTERISTICS:

Query Match 12.6%, Score 67.5; DB 12; Length 1404;
Best Local Similarity 23.6%, Pred. No. 95;
Matches 21; Conservative 17; Mismatches 24; Indels 27; Gaps 3;

QY 19 APSTLRVTMAEFRVPLPP-----MAEVPMLSTQNSMGSSASASASSLEMW 65
Db 125 ASQIKSTT-----KRSPKPPNKKTKKVEIEEETEEHSVSENQESSSSSSSSSTI 180

QY 66 E-----KDLEERLNSIDHDMNNK 84
Db 181 KIKSKNSAANRELQKCLKVKDNKKNRTK 209

RESULT 14

US-10-124-557-62

; Sequence 62, Application US/10124557
; Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 62:
SEQUENCE DESCRIPTION: SEQ ID NO: 62:
MOLECULE TYPE: protein
LENGTH: 1404 amino acids
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE CHARACTERISTICS:

Query Match 12.6%, Score 67.5; DB 12; Length 1404;
Best Local Similarity 23.6%, Pred. No. 95;
Matches 21; Conservative 17; Mismatches 24; Indels 27; Gaps 3;

QY 19 APSTLRVTMAEFRVPLPP-----MAEVPMLSTQNSMGSSASASASSLEMW 65
Db 125 ASQIKSTT-----KRSPKPPNKKTKKVEIEEETEEHSVSENQESSSSSSSSSTI 180

QY 66 E-----KDLEERLNSIDHDMNNK 84
Db 181 KIKSKNSAANRELQKCLKVKDNKKNRTK 209

RESULT 15

US-09-835-996A-17

; Sequence 17, Application US/09835996A

; Patent No. US20020142953A1

GENERAL INFORMATION:

APPLICANT: Ballinger, Dennis

APPLICANT: Loeb, Debra

APPLICANT: Montgomery, Julie

APPLICANT: Tang, Y. Tom

APPLICANT: Zhou, Ping

APPLICANT: Goodrich, Ryle

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhao, Qing

APPLICANT: Wehrman, Tom

APPLICANT: Drmanac, Radoje

APPLICANT: Ren, Feiyang

APPLICANT: Qian, Xiaohong

APPLICANT: Wang, Dunrui

TITLE OF INVENTION: MATERIALS AND METHODS RELATING TO LIPID METABOLISM

FILE REFERENCE: 28110/35915A

CURRENT APPLICATION NUMBER: US/09/835,996A

CURRENT FILING DATE: 2001-04-16

PRIOR APPLICATION NUMBER: US 60/197,137

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: US 09/714,936

PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: US 09/667,298

PRIOR FILING DATE: 2000-09-22

PRIOR APPLICATION NUMBER: US 09/631,451

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: US 09/598,042

PRIOR FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 45

SOFTWARE: PatentIn version 3.0

```
; SEQ ID NO 17
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-835-996A-17
```

```
Query Match      12.5%; Score 67; DB 10; Length 348;
Best Local Similarity 45.2%; Pred. NO. 17;
Matches 14; Conservative 4; Mismatches 13; Indels 0; Gaps 0;
```

```
Qy 56 SASASSLEMEKLEERLNSIDHDMNNKFG 86
Db 318 SESAESLRQWAELEENLNELTHIQSLKAG 348
```

```
Search completed: July 16, 2003, 12:51:10
Job time : 51 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:45:51 ; Search time 11 Seconds
(without alignments)
392.140 Million cell updates/sec

Title: US-09-814-661A-2
Perfect score: 104
Sequence: 1 MNSQDYFYAQNRCQQQAP.....FGSGELKSMFNGKVEEMDF 104

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	104	1 YMF8_YEAST	Q04964 saccharomyc
2	8	7.7	274	1 PSBS_TOBAC	Q9SMB4 nicotiana t
3	8	7.7	317	1 WRIS_ARATH	O22176 arabidopsis
4	8	7.7	411	1 CRFL_YARLI	P45815 yarrowia li
5	8	7.7	474	1 SOS4_HUMAN	Q06945 homo sapien
6	8	7.7	492	1 SES3_MOUSE	O9CYP7 mus musculu
7	8	7.7	522	1 RECN_HELPJ	O9ZJ80 helicobacte
8	8	7.7	523	1 C756_CAME	O04773 campanula m
9	8	7.7	524	1 RECN_HELPY	O25943 helicobacte
10	8	7.7	670	1 DOT6_YEAST	P40059 saccharomyc
11	8	7.7	787	1 MADE_DROME	O9VYA0 drosophila
12	7	6.7	203	1 YH14_YEAST	P38897 saccharomyc
13	7	6.7	229	1 RNC_PSEAE	O9XCX9 pseudomonas
14	7	6.7	332	1 YFEH_ECOLI	P39836 escherichia
15	7	6.7	334	1 ROC2_ARATH	Q43349 arabidopsis
16	7	6.7	369	1 MAF_RAT	P54844 rattus norv
17	7	6.7	396	1 UGAT_HUMAN	P78381 homo sapien
18	7	6.7	539	1 PNUT_DROME	P40797 drosophila
19	7	6.7	568	1 DISC_DROME	P23792 drosophila
20	7	6.7	590	1 HMAA_DROME	P29555 drosophila
21	7	6.7	606	1 MGD2_HUMAN	Q9UNF1 homo sapien
22	7	6.7	611	1 VIE3_MCMVS	P29832 murine cyto
23	7	6.7	664	1 PLB1_YEAST	P39105 saccharomyc
24	7	6.7	673	1 SIM_DROME	P05709 drosophila
25	7	6.7	760	1 RESA_PLAFN	P13831 plasmodium
26	7	6.7	819	1 ADVL_MOUSE	O88398 mus musculu
27	7	6.7	822	1 YJFL_YEAST	P47046 saccharomyc
28	7	6.7	939	1 ST20_YEAST	Q03497 saccharomyc
29	7	6.7	958	1 AMYG_DEBOC	P22861 debaryomyc
30	7	6.7	1073	1 RESA_PLAFF	P13830 plasmodium
31	7	6.7	1391	1 N155_HUMAN	O75694 homo sapien
32	7	6.7	1391	1 N155_MOUSE	Q99P88 mus musculu
33	7	6.7	3726	1 TRX_DROME	P20659 drosophila

34 7 6.7 3828 1 TRX_DROVI
35 6 5.8 27 1 L52_ADE07
36 6 5.8 40 1 PHAC_MASLA
37 6 5.8 59 1 SECE_TREPA
38 6 5.8 72 1 RS18_FUSNN
39 6 5.8 107 1 TBCA_BOVIN
40 6 5.8 107 1 TBCA_HUMAN
41 6 5.8 107 1 TBCA_MOUSE
42 6 5.8 107 1 TBCA_RABIT
43 6 5.8 108 1 KVIS_HUMAN
44 6 5.8 115 1 VS_BPT4
45 6 5.8 117 1 PT26_STYPL

Q24742 drosophila
P05663 human adeno
P11389 mastigoclad
O83263 treponema p
O8rie4 fusobacteri
P48427 bos taurus
O75347 homo sapien
P48428 mus musculu
P80584 oryctolagus
P01611 homo sapien
P13310 bacteriophag
P28218 styela plic

ALIGNMENTS

RESULT 1
YMF8_YEAST
ID YMF8_YEAST STANDARD; PRT; 104 AA.
AC 004964;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 11.8 kDa protein in OG1-CNA2 intergenic region.
GN YML058W OR YM9958.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Devlin K, Churcher C, Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46729; CAA86717.1;
DR SGD; S0004523; YML058W.
KW Hypothetical protein
SQ SEQUENCE 104 AA; 11834 MW; 02735C9178E431B2-CRC64;
Query Match 100.0%; Score 104; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 5.5e-100;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSQDYFYAQNRCQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSMGSSASAS 60
|||||
DB 1 MNSQDYFYAQNRCQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSMGSSASAS 60
|||||
QY 61 SLEWKEKDLERLNSIDHDMMNNKFGSGELKSMFNGKVEEMDF 104
|||||
DB 61 SLEWKEKDLERLNSIDHDMMNNKFGSGELKSMFNGKVEEMDF 104
|||||

RESULT 2
PSBS_TOBAC
ID PSBS_TOBAC STANDARD; PRT; 274 AA.
AC Q9SMB4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 22 kDa protein, chloroplast precursor (CP22).
GN PSBS
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch)

CC -----
 DR EMBL: Z23265; CAA80803.1; -
 DR HSP: P41772; LCO4
 DR InterPro: IPR001083; Copper-fist.
 DR Pfam: PF00649; Copper-fist; 1.
 DR PRINTS: PR00617; COPPERFIST.
 DR ProDom: PD009050; Copper-fist; 1.
 DR SMART: SM00412; Cu.FIST; 1.
 DR PROSITE: PS01119; COPPER_FIST_1; 1.
 DR PROSITE: PS00073; COPPER_FIST_2; 1.
 KW Transcription regulation; DNA-binding; Copper; Nuclear protein.
 FT DNA_BIND 1 40
 FT DOMAIN 63 66 COPPER-FIST.
 FT DOMAIN 126 140 POLY-GLN.
 FT DOMAIN 205 211 POLY-SER.
 FT METAL 11 11 ZINC (BY SIMILARITY).
 FT METAL 14 14 ZINC (BY SIMILARITY).
 FT METAL 23 23 ZINC (BY SIMILARITY).
 FT METAL 25 25 ZINC (BY SIMILARITY).
 SQ SEQUENCE 411 AA; 43688 MW; 9AD602FA467ERCC7 CRC64;

Query Match 7.7%; Score 8; DB 1; Length 411;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
 Db 356 SASASASS 363

RESULT 5

SOX4_HUMAN
 ID SOX4_HUMAN STANDARD; PRT; 474 AA.
 AC Q06945;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription factor SOX-4.
 GN SOX4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94093204; PubMed=8258656;
 RA Farr C.J., Easty D.J., Ragoussis J., Collignon J., Lovell-Badge R.,
 RA Goodfellow P.N.;
 RT "Characterization and mapping of the human SOX4 gene."
 RL Mamm. Genome 4:577-584(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Smalley C.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 70-123 FROM N.A.
 RX MEDLINE=92310993; PubMed=1614875;
 RA Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;
 RT "A conserved family of genes related to the testis determining gene,
 RT SRY".
 RL Nucleic Acids Res. 20:2887-2887(1992).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS WITH HIGH AFFINITY
 CC TO THE T-CELL ENHANCER MOTIF 5'-ACAAAG-3' MOTIF.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: TESTIS, BRAIN, AND HEART.
 CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch)

CC -----
 DR EMBL: X70683; CAA50018.1; -
 DR DR EMBL: AL136179; CAC03595.1; -
 DR EMBL: X65661; CAA46612.1; -
 DR PIR: S21479; S21479.
 DR PIR: S22938; S22938.
 DR PIR: S31724; S31724.
 DR HSP: P48436; ISX9.
 DR TRANSFAC: T01276; -
 DR Genew: HGNC:11200; SOX4.
 DR MIM: 184430; -
 DR InterPro: IPR000910; HMG_12_box.
 DR Pfam: PF00505; HMG_box; 1.
 DR SMART: SM00398; HMG; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
 FT DNA_BIND 59 127 HMG BOX.
 FT DOMAIN 386 397 POLY-SER.
 FT CONFLICT 71 71 Q -> P (IN REF. 3).
 SQ SEQUENCE 474 AA; 47262 MW; CB75B76ABF507A07 CRC64;

Query Match 7.7%; Score 8; DB 1; Length 474;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
 Db 266 SASASASS 273

RESULT 6

SES3_MOUSE
 ID SES3_MOUSE STANDARD; PRT; 492 AA.
 AC Q9CYP7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sestrin 3.
 GN SES3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6J; TISSUE=Embryo;
 MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinctsch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: BELONGS TO THE SESTRIN FAMILY.
 CC -----
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AK017464; BAB30755.1; -
 DR MGD; MGI:1922997; 5630400E15R1K.
 SQ SEQUENCE 492 AA; 57020 MW; 8E9AD9CB45656827 CRC64;

Query Match 7.7%; Score 8; DB 1; Length 492;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GSSASASA 59
 |||||
 DB 5 GSSASASA 12

RESULT 7
 RECN_HELPJ STANDARD; PRT; 522 AA.
 ID RECN_HELPJ
 AC Q9ZJ80;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA repair protein recN (Recombination protein N).
 GN RECN OR JHP1434.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voris G.F.,
 RA Trust T.J.;
 RT Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.;
 RL Nature 397:176-180(1999).
 CC -!- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
 CC DNA (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE RECN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE001566; AAD07019.1; -
 DR InterPro: IPR003439; ABC_transport.
 KW DNA repair; ATP-binding; Complete proteome.
 FT NP_BIND 36 43 ATP (POTENTIAL).
 SQ SEQUENCE 522 AA; 59349 MW; 000D7EE3C0632EA5 CRC64;

Query Match 7.7%; Score 8; DB 1; Length 522;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KLEERLN 74
 |||||
 DB 185 KLEERLN 192

RESULT 8

C756 CAMME STANDARD; PRT; 523 AA.
 ID C04773;
 AC 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Flavonoid 3',5'-hydroxylase (EC 1.14.-.-) (E3'5'H) (Cytochrome P450
 75A6).

GN CYP75A6.
 OS Campanula medium (Canterbury bells).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids II; Asterales; Campanulaceae; Campanula.

OX NCBI_TaxID=56154;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Petal;
 RA Ohbayashi M.;

RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CATALYSES THE 3'5'-HYDROXYLATION OF NARINGENIN AND
 CC ERIODICTYOL TO FORM 5,7,3',4',5'-PENTAHYDROXYFLAVANONE AND 3',5'-
 CC HYDROXYLATION OF DIHYDROKAEEMPEROL AND DIHYDROQUERCETIN TO FORM
 CC DIHYDROMYRICETIN (BY SIMILARITY).
 CC -!- PATHWAY: Anthocyanin biosynthesis.
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D14590; BAA03440.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; p450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 460 460 HEME (BY SIMILARITY).
 SQ SEQUENCE 523 AA; 58726 MW; 70E39C6B76F387D9 CRC64;

Query Match 7.7%; Score 8; DB 1; Length 523;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
 |||||
 DB 209 SSASASAS 216

RESULT 9
 RECN_HELPJ STANDARD; PRT; 524 AA.
 ID RECN_HELPJ
 AC O25943;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA repair protein recN (Recombination protein N).
 GN RECN OR HPI393.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori*.";
 RL Nature 388:539-547(1997).
 CC -!- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
 CC DNA (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE REC N FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000639; AAD08433.1; -;
 DR TIGR; HP1393; -;
 DR InterPro; IPR003439; ABC transportr.
 KW DNA repair; ATP-binding; Complete proteome.
 FT NP_BIND 36 43 ATP (POTENTIAL).
 SQ SEQUENCE 524 AA; 59564 MW; 1F78BFF0CFCE568A CRC64;
 Query Match 7.7%; Score 8; DB 1; Length 524;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 67 KDLERLN 74
 DB 185 KDLERLN 192
 RESULT 10
 DOT6_YEAST
 ID DOT6_YEAST STANDARD; PRT; 670 AA.
 AC P40059;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Disrupter of telomere silencing protein 6.
 GN DOT6 OR YER088C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RX MEDLINE=98429517; PubMed=9755194;
 RA Singer M.S., Kahana A., Wolf A.J., Meisinger L.L., Peterson S.E.,
 RA Goggin C., Mahowald M., Gottschling D.E.,
 RT "Identification of high-copy disruptors of telomeric silencing in
 RT *Saccharomyces cerevisiae*.";
 RL Genetics 150:613-632(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Has a role in telomere silencing, which is the
 CC repression of chromatin structure which leads to a stop in the
 CC transcription of nearby genes.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U18839; AAB64643.1; -;
 DR SGD; S0000890; DOT6.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR Pfam; PF00249; myb_DNA-binding; 1.
 DR SMART; SM00395; SANT; 1.
 DR PROSITE; PS00037; MYB.1; 1.
 DR PROSITE; PS00334; MYB.2; FALSE_NEG.
 DR PROSITE; PS00090; MYB.3; 1.
 KW Transcription regulation; Nuclear protein; DNA-binding.
 FT DNA_BIND 67 117 MYB.
 SQ SEQUENCE 670 AA; 73048 MW; 61FD64749653FE78 CRC64;
 Query Match 7.7%; Score 8; DB 1; Length 670;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 54 SASASASS 61
 DB 58 SASASASS 65
 RESULT 11
 NADE_DROME
 ID NADE_DROME STANDARD; PRT; 787 AA.
 AC Q9VVA0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative glutamine-dependent NAD(+) synthetase (EC 6.3.5.1) (NAD(+)
 DE synthase [glutamine-hydrolyzing]).
 GN CG9940.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Muzny L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -1- CATALYTIC ACTIVITY: ATP(+) + L-glutamine + H(2)O -
CC AMP + diphosphate + NAD(+) + L-glutamate.
CC -1- PATHWAY: NAD biosynthesis.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE NAD
CC SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF003493; AAF48303.1; -
DR FlyBase: FBgn0030512; CG9940.
DR InterPro: IPR003694; NAD_synthase.
DR InterPro: IPR003010; Ntlase/CNhydase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR Pfam: PF02540; NAD_synthase; 1.
DR TIGRFAMs: TIGR00552; nadE; 1.
KW Hypothetical protein; Ligase; NAD; ATP-binding.
FT DOMAIN 325 787
FT NP_BIND 355 362 ATP (BY SIMILARITY).
FT ACT_SITE 357 357 BY SIMILARITY.
FT SEQUENCE 787 AA; 87614 MW; 0B2B6106E327F0C5 CRC64;
SQ

Query Match 7.7%; Score 8; DB 1; Length 787;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 SSASASAS 60
Db |||||
733 SSASASAS 740

RESULT 12
YH14_YEAST
ID YH14_YEAST STANDARD; PRT; 203 AA.
AC P38997;
RA 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 20.6 kDa protein in FLO5-PH012 intergenic region.
GN YHR214W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA Vaudin M.;

RT "Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome
VIII.";
RL Science 265:2077-2082(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00029; AAB69739.1; -
DR PIR: S48995; S48995.
DR SGD: S0001257; YHR214W.
KW Hypothetical protein.
SQ SEQUENCE 203 AA; 20593 MW; 466EC45BECC69C CRC64;
Query Match 6.7%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 SSASASA 59
Db |||||
42 SSASASA 48

RESULT 13
RNC_PSEAE
ID RNC_PSEAE STANDARD; PRT; 229 AA.
AC Q9XCX9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease III (EC 3.1.26.3) (RNase III).
GN RNC OR PA0770.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK;
RX MEDLINE=99369891; PubMed=10438789;
RA Powell B.S., Peters H.K. III, Nakamura Y., Court D.L.;
RT "Cloning and analysis of the rnc-*era*-*recO* operon from *Pseudomonas*
RT *aeruginosa*.";
RL J. Bacteriol. 181:5111-5113(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AF123492; RAD40229.1; -;
 DR EMBL: AE004512; AG04159.1; -;
 DR HSSP: Q91836; IDI2.
 DR InterPro: IPR001159; DS_RBD.
 DR InterPro: IPR000999; RNase_3.
 DR Pfam: PF00035; dsrm; 1.
 DR Pfam: PF00636; Ribonuclease_3; 1.
 DR SMART: SM00358; DSRM; 1.
 DR SMART: SM00355; RiboQ; 1.
 DR PROSITE: PS00137; DS_RBD; 1.
 DR PROSITE: PS00517; RNase_3.1; 1.
 DR PROSITE: PS0142; RNase_3.2; 1.
 DR Hydrolase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
 FT DOMAIN 5 127 RNase III.
 FT DOMAIN 154 224 DSRM.
 FT SEQUENCE 229 AA; 23506 MW; 9D884F21A008B6E CRC64;

Query Match 6.7%; Score 7; DB 1; Length 229;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 GSGELKS 92
 Db 96 GSGELKS 102

RESULT 14

YFEH_ECOLI
 ID YFEH_ECOLI STANDARD; PRT; 332 AA.
 AC P39836; P76532; P76958;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yfeh.
 GN YFEH OR B2410.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 K-12 genome corresponding to 50.0-68.8 min on the linkage map and
 analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [3]
 RP SEQUENCE OF 94-332 FROM N.A.
 RC O'Connor M.J., Ally A., Ally D., Zhang X., Robichaud M., Backman K.;
 RX Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP IDENTIFICATION.

RX MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: TO P.AERUGINOSA HYPOTHETICAL PROTEIN IN GOR 3'
 CC REGION (AC P39879).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000328; AAC75463.1; -;
 DR EMBL: D90870; BAA16281.1; ALT_INIT.
 DR EMBL: M24278; -; NOT_ANNOTATED_CDS.
 DR EcoGene; EG12376; yfeh.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 325 332 QESSADKA -> RKAAPIKLKADASGAGPVASAPRF (IN
 FT REF. 3).
 SQ SEQUENCE 332 AA; 36405 MW; B8C96697C5BC30F1 CRC64;

Query Match 6.7%; Score 7; DB 1; Length 332;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 SASASSL 62
 Db 132 SASASSL 138

RESULT 15
 ROC2_ARATH
 ID ROC2_ARATH STANDARD; PRT; 334 AA.
 AC Q43349; Q9LFH2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 29 kDa ribonucleoprotein, chloroplast precursor (RNA-binding protein
 DE cp29).
 GN RBP29 OR AT3C53460 OR F4P12.160.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Leaf;
 RX MEDLINE=95201246; PubMed=7894017;
 RA Ohta M., Sugita M., Sugiyama M.;
 RT "Three types of nuclear genes encoding chloroplast RNA-binding
 RT proteins (cp29, cp31 and cp33) are present in Arabidopsis thaliana:
 RT presence of cp31 in chloroplasts and its homologue in
 RT nuclei/cytoplasm.";
 RL Plant Mol. Biol. 27:529-539(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
 RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Deisen J., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordtsiek G.,

RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laude M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana";
RL Nature 408:820-822(2000).
CC -!- FUNCTION: COULD BE INVOLVED IN SPLICING AND/OR PROCESSING OF
CC CHLOROPLAST RNA'S.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D31710; BAA06518.1; -;
DR EMBL: D31711; BAA06519.1; -;
DR EMBL: AL132966; CAB67653.1; ALT_SEQ.
DR HSSP: P09651; IHA1.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS50102; RRM; 2.
DR RNA-binding: Ribonucleoprotein; Repeat; mRNA processing; Chloroplast;
KW RNA-binding; Ribonucleoprotein; Repeat; mRNA processing; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 65 CHLOROPLAST (POTENTIAL).
FT CHAIN 66 334 29 KDA RIBONUCLEOPROTEIN.
FT DOMAIN 99 177 RNA-BINDING (RRM) 1.
FT DOMAIN 178 248 LINKER (GLY-RICH).
FT DOMAIN 249 327 RNA-BINDING (RRM) 2.
SQ SEQUENCE 334 AA; 35243 MW; B117D601790C07A CRC64;

Query Match 6.7%; Score 7; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 9,9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SASASL 62
| | | | |
Db 2 SASASL 8

Search completed: July 16, 2003, 12:52:10
Job time : 12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:25:36 ; Search time 70 Seconds
(without alignments)
197.972 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 537

Sequence: 1 MONSDYFYAQNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	537	100.0	104	21	Yeast Sm11 (suppre
2	82	15.3	1665	22	Drosophila melanog
3	80	14.9	1982	22	Novel human diagno
4	80	14.9	1982	22	Novel human diagno
5	75	14.0	560	22	Amino acid sequenc
6	73	13.6	1028	22	Drosophila melanog
7	71.5	13.3	328	22	Drosophila melanog
8	71	13.2	372	22	Drosophila melanog
9	71	13.2	810	20	Chlamydia pneumoni
10	71	13.2	811	23	Chlamydia pneumoni

11	71	13.2	1833	22	ABB71141	Drosophila melanog
12	70.5	13.1	166	22	AAG61796	Interferon alpha h
13	70	13.0	176	22	ABB62448	Drosophila melanog
14	70	13.0	624	22	ABB70398	Drosophila melanog
15	69	12.8	1063	23	ABP35715	Fungal ZBC protein
16	68.5	12.8	136	21	AAG35938	Zea mays protein f
17	68.5	12.8	166	22	AAG61819	Interferon alpha h
18	68.5	12.8	334	22	ABG08066	Novel human diagno
19	68.5	12.8	1068	23	ABP35714	Fungal ZBC protein
20	68	12.7	1019	22	ABB63588	Drosophila melanog
21	68	12.7	1019	22	AAU38953	Drosophila G-prote
22	68	12.7	2742	21	AAAB23012	Human APC protein
23	68	12.7	2842	15	AAR63508	Adenomatous polypo
24	68	12.7	2843	13	AAR26052	APC gene product i
25	68	12.7	2843	16	AAW11922	Adenomatous polypo
26	68	12.7	2843	18	AAW35392	Human adenomatous
27	68	12.7	2843	19	AAW76140	Human APC protein
28	68	12.7	2843	19	AAW76144	Human adenomatous
29	68	12.7	2843	19	AAW38370	Human APC protein
30	68	12.7	2843	21	AAAB23011	Human APC protein
31	68	12.7	2860	15	AAR63507	Adenomatous polypo
32	68	12.7	2973	19	AAW76821	Human APC protein.
33	68	12.7	2973	21	AAW70304	Protein used in ca
34	68	12.7	2973	22	ABG20855	Transcriptional ac
35	67.5	12.6	159	22	ABG20855	Novel human diagno
36	67.5	12.6	336	21	AAG29574	Arabidopsis thalia
37	67.5	12.6	397	22	ABG68819	Drosophila melanog
38	67.5	12.6	445	21	AAAG29573	Arabidopsis thalia
39	67.5	12.6	493	22	AAAB31432	Amino acid sequenc
40	67.5	12.6	571	21	AAAG29572	Arabidopsis thalia
41	67.5	12.6	1299	22	AAW24322	Human EST encoded
42	67.5	12.6	1300	22	ABG64083	Drosophila melanog
43	67.5	12.6	1404	13	AAR26049	MSF precursor. Sy
44	67.5	12.6	1404	22	AAAB60568	Human megakaryocyt
45	67.5	12.6	1404	22	AAAB29773	Human megakaryocyt

ALIGNMENTS

RESULT 1
AA90470
ID AA90470 standard; Protein: 104 AA.
XX
AC AA90470;
XX
DT 15-AUG-2000 (first entry)
XX
DE Yeast Sm11 (suppressor of mecl lethality) protein.
XX
KW Sm11 protein; YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200017225-A2.
XX
PD 30-MAR-2000.
XX
PF 24-SEP-1999; 99WO-US22260.
XX
PR 24-SEP-1998; 98US-0158858.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Rothstein R, Zhao X;
XX
DR WPI; 2000-283539/24.
XX
N-PSDB; AAA14263.

PT New Smll protein and its homologs, useful for treating cancer,
PT microbial infection and ataxia telangiectasia and in screening for
PT specific modulators
XX
XX Claim 2: Fig 1C; 98pp; English.
XX This sequence represents the yeast Smll (suppressor of mecl lethality)
CC protein. This protein is encoded by the SML1 gene, located on chromosome
CC XIII, and is an inhibitor of ribonucleotide reductase. The SML1 gene
CC (assigned the name YML058w by the Stanford yeast genomic project) was
CC initially isolated in a mecl-1 strain by genetic analysis. Smll is a
CC suppressor of the mecl mutant in yeast which is associated with abnormal
CC levels of recombination in both meiosis and mitosis. Smll also permits
CC cell growth in the absence of the cell cycle checkpoint proteins Mecl
CC and Rad53, which, unlike most checkpoint proteins, are essential for cell
CC growth. Smll binds to the largest subunit of ribonucleotide reductase
CC (RNR) which catalyses the rate-limiting step of deoxynucleotide
CC triphosphate (dNTP) synthesis, conversion of NTPs to dNTPs. Binding of
CC Smll reduces the activity of RNR and thus inhibits the synthesis of
CC dNTPs. The SML1 gene, or fragments thereof may be used to isolate the
CC human homologue of SML1. The Smll protein, and compounds that modulate
CC the interaction of Smll with ribonucleotide reductase (RNR), may be used
CC to alter the rate at which cells divide. These are particularly useful
CC for treating cancer, microbial infection and ataxia telangiectasia (AT),
CC a condition in which patients are prone to cancer. Smll may also be used
CC to screen for modulatory agents, to raise specific antibodies, and for
CC stimulating the function of the ATM gene (a mammalian Mecl homologue
CC which is mutated in AT). Anti-Smll antibodies are used as diagnostic and
CC analytical immunoassay reagents and to remove Smll from serum or to
CC titrate Smll intracellularly.

SQ Sequence 104 AA;

Query Match 100.0%; Score 537; DB 21; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.9e-58;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSODYFYAQRCCQQQAPSTLRVTMAEFRVRPLPPMAEVPMLSTQNSMGSSASAS 60
DB 1 MNSODYFYAQRCCQQQAPSTLRVTMAEFRVRPLPPMAEVPMLSTQNSMGSSASAS 60
QY 61 SLEWEKDLERLNSIDHDMNNKFGSGELKSMFNQGVKVEEMDF 104
DB 61 SLEWEKDLERLNSIDHDMNNKFGSGELKSMFNQGVKVEEMDF 104

RESULT 2

ABB64010
ID ABB64010 standard; Protein; 1665 AA.

AC ABB64010;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 18822.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
DR N-PSDB; ABL08113.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX Disclosure; SEQ ID NO 18822; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1665 AA;

Query Match 15.3%; Score 82; DB 22; Length 1665;
Best Local Similarity 28.6%; Pred. No. 2.2;
Matches 28; Conservative 15; Mismatches 41; Indels 14; Gaps 3;

QY 10 AQNRCCQQQAPSTLRVTMAEFRVRPLPPMAEVPMLSTQNSMGSSAS--ASASSLEMWK 67
DB 894 ARKREHSSQSSNGNTPT---KKVATPOLVAAPLKPTNTAGSSSSDESSSAESSK 949
QY 68 DLEERLNSIDHDMNN-----NKFGSGELKSMFNQGV 97
DB 950 SSSSSSSDDTETQNTNCRIVKLNKTGAVQKALLGSG 987

RESULT 3

ABG16404
ID ABG16404 standard; Protein; 1982 AA.

XX ABG16404;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #16395.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS80591.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

Claim 20: SEQ ID No 46763; 103pp: English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX		New isolated polynucleotide and encoded polypeptides, useful in
PT		diagnostics, forensics, gene mapping, identification of mutations
PT		responsible for genetic disorders or other traits and to assess
PT		biodiversity
XX		
PS	Claim 20; SEQ ID No 50015; 103pp; English.	
XX		The invention relates to isolated polynucleotide (I) and
CC		polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC		polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC		and gene mapping, and in recombinant production of (II). The
CC		polynucleotides are also used in diagnostics as expressed sequence tags
CC		for identifying expressed genes. (I) is useful in gene therapy techniques
CC		to restore normal activity of (II) or to treat disease states involving
CC		(II). (II) is useful for generating antibodies against it, detecting or
CC		quantitating a polypeptide in tissue, as molecular weight markers and as
CC		a food supplement (II) and its binding partners are useful in medical
CC		imaging of sites expressing (II). (I) and (II) are useful for treating
CC		disorders involving aberrant protein expression or biological activity.
CC		The polypeptide and polynucleotide sequences have applications in
CC		diagnostics, forensics, gene mapping, identification of mutations
CC		responsible for genetic disorders or other traits to assess biodiversity
CC		and to produce other types of data and products dependent on DNA and
CC		amino acid sequences. ABG00010-ABG30377 represent novel human
CC		diagnostic amino acid sequences of the invention.
CC		Note: The sequence data for this patent did not appear in the printed
CC		specification, but was obtained in electronic format directly from WIPO
CC		at ftp.wipo.int/pub/published_pct_sequences .
XX		
SQ	Sequence 1982 AA;	
	Query Match 14.9%; Score 80; DB 22; Length 1982;	
	Best Local Similarity 33.9%; Pred. No. 5;	
	Matches 41; Conservative 13; Mismatches 35; Indels 32; Gaps 9;	
QY	2 QNSGDVFFYAQNRCQQQAAPS--TLRTVTMAEF-----RRVPLPPMAEVPMLSTQNSMGSSA 55	
DB	::::: :: :: :: :: :: :: :: :: ::	
DB	1707 ENKSDVTAAGNLVRESGASPITFTSLREAEFSDNORRLS-PPF---PLEPAQKRVRSSP 1762	
QY	56 SA-----SASSLEMKWD- ---LLERLASID-HDMNNKFSGSELKSMFNCKGYEEMD 103	
DB	1763 LASFLOQRSAASLE-WEPEPHLYRSKLKLSINVH-----GDLRLKSHPKVKVERH 1812	
QY	104 F 104	
DB	1813 F 1813	
RESULT 5		
AAG67397		
ID	AAG67397 standard; Protein; 560 AA.	
AC	AAG67397;	
XX		
DT	13-NOV-2001 (first entry)	
XX		
DE	Amino acid sequence of human protein kinase SGK424.	
KW	Human; protein kinase; cancer; immune disease; cardiovascular disease;	
KW	brain disease; neuronal disease; Alzheimer's disease; chromosome 19;	
KW	Parkinson's disease; multiple sclerosis; metabolic disorder;	
KW	peripheral nervous system disease; amyotrophic lateral sclerosis;	
KW	infection; ocular disease; migraine; pain; sexual dysfunction;	
KW	mood disorder; attention disorder; cognition disorder; hypotension;	
KW	hypertension; psychotic disorder; dyskinesia; transplant rejection.	
XX		
OS	Homo sapiens.	
XX		
PN	WC200166594-A2.	
XX		
PD	13-SEP-2001.	
XX		

```
PF 02-MAR-2001; 2001WO-US06838.
XX
XX
PR 06-MAR-2000; 2000US-0187150.
PR 29-MAR-2000; 2000US-0193404.
PR 13-NOV-2000; 2000US-0247013.
XX
XX (SUGE-) SUGEN INC.
PA
XX
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI
XX
XX WPI: 2001-536777/59.
DR N-PSDB; AAH77996.
XX
XX Nucleic acids capable of encoding human polypeptides having a kinase or
PT kinase-like activity, useful for diagnosing a disease selected from
PT cancers, cardiovascular disease and neuronal-associated diseases (e.g.
PT Alzheimer's disease)
XX
XX Claim 6; Fig 2C; 20lpp; English.
XX
XX The present sequence represents a human protein kinase. The gene is
CC located at chromosomal position 19q12-q13.3. The kinase polypeptides
CC are useful for diagnosing a disease or disorder selected from cancers
CC (e.g. cancers of tissues and cancers of hematopoietic origin),
CC immune-related diseases and disorders, cardiovascular disease, brain
CC or neuronal-associated diseases (e.g. Alzheimer's disease, Parkinson's
CC disease, multiple sclerosis), metabolic disorders, peripheral nervous
CC system diseases, amyotrophic lateral sclerosis, viral infections,
CC infections caused by prions, infections caused by bacteria, infections
CC caused by fungi, ocular diseases, migraines, pain, sexual dysfunction,
CC mood disorders, attention disorders, cognition disorders, hypotension,
CC hypertension, psychotic disorders, dyskinesias, and organ transplant
CC rejection. Kinase inhibitors are useful for treating diseases and
CC disorders described above.
XX
XX
SQ Sequence 560 AA;
Query Match 14.0%; Score 75; DB 22; Length 560;
Best Local Similarity 35.9%; Pred. NO. 3.4;
Matches 23; Conservative 7; Mismatches 32; Indels 2; Gaps 1;
QY 36 LPPMAEVPMLSTQNSMGSSASASASLEMKWDLERLNSIDHDMNNKFGSGELKSMFN 95
DB 190 LPPMASLDQL--QAREGNSPPGSLSLRLMQADRLRLAQIRRASQVPALGFGSLSLWP 247
QY 96 QGKV 99
DB 248 PGLV 251
RESULT 6
ABB62462
ID ABB62462 standard; Protein; 1028 AA.
AC
AC ABB62462;
XX
XX 26-MAR-2002 (first entry)
DT
DE Drosophila melanogaster polypeptide SEQ ID NO 14178.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX WPI: 2001-536777/59.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PR
```

```
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI: 2001-656860/75.
DR N-PSDB; ABL06565.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
XX Disclosure; SEQ ID NO 14178; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72073).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1028 AA;
Query Match 13.6%; Score 73; DB 22; Length 1028;
Best Local Similarity 21.4%; Pred. NO. 14;
Matches 25; Conservative 17; Mismatches 33; Indels 42; Gaps 2;
QY 27 TMAEFRRVPLP-----PMAFVPMPLSTQNSMGSSASASASLEMM 65
DB 397 SVLEFPKPEPMNEFNWDFEKNFYDYIKEENPLQEMPMPIKENSKENVDASSVDLENL 456
QY 66 ERD-----LEERLNSIDHDMNNKFGSGELKSMFNQGVKEE 101
DB 457 QKDQDGRDRDPKLPALENVETQKELELENLPESHNNKRLRVKEKEIVROGKLKE 513
RESULT 7
ABB69217
ID ABB69217 standard; Protein; 328 AA.
XX
AC ABB69217;
XX
XX 26-MAR-2002 (first entry)
DT
DE Drosophila melanogaster polypeptide SEQ ID NO 34443.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI: 2001-656860/75.
DR N-PSDB; ABL13320.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT
```


Best Local Similarity 27.8%; Pred. No. 18;
Matches 22; Conservative 11; Mismatches 20; Indels 26; Gaps 2;

QY 14 CQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSGSSASASASLEWWEK----- 67
Db 559 CRIEIEKTLR-----MAELPLLTKKAFKACSQYNCAEMLEKVKPYCK 604
QY 68 -----DLEERLNSIDHDM 80
Db 605 ESLAYVTSKERLVSLEDEL 623

RESULT 10
ABB90703
ID ABB90703 standard; Protein; 811 AA.

AC ABB90703;
DT 29-JUL-2002 (first entry)
XX Chlamydia pneumoniae cp7388 protein, SEQ ID NO:355.

DE Chlamydia pneumoniae; antigen; immunogen; vaccine; diagnosis;
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029.

OS Chlamydia pneumoniae.

XX WO2000202606-A2.

PN 10-JAN-2002.

PD 03-JUL-2001; 2001WO-IB01445.

PE 03-JUL-2000; 2000GB-0016363.

PR 11-JUL-2000; 2000GB-0017047.

PR 21-JUL-2000; 2000GB-0017983.

PR 07-AUG-2000; 2000GB-0019368.

PR 18-AUG-2000; 2000GB-0020440.

PR 14-SEP-2000; 2000GB-0022583.

PR 10-NOV-2000; 2000GB-0027549.

PR 22-DEC-2000; 2000GB-0031706.

XX (CHIR-) CHIRON SPA.

PI Ratti G, Grandi G;

XX WPI; 2002-154726/20.

DR N-PSDB; ABL91361.

XX Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT medicament for treatment or prevention of infection due to Chlamydia,
PT preferably Chlamydia pneumoniae, and for diagnostic purposes -
XX Claim 1; Page 179; 364pp; English.

XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC them. The proteins are predicted to be immunogenic and may therefore be
CC useful in vaccine production and for diagnostic purposes. Chlamydia
CC pneumoniae is a common cause of respiratory disease in humans, and is
CC also involved in the development of cardiovascular diseases such as
CC atherosclerosis, coronary artery disease, carotid artery stenosis,
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC claudication and stroke. The proteins and nucleic acids of the invention
CC may be used in vaccines and pharmaceutical compositions for the
CC prevention or treatment of chlamydial infections, particularly Chlamydia
CC pneumoniae infections. The proteins may also be used in the detection of
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a

CC specifically claimed Chlamydia pneumoniae protein of the invention.
XX
SQ Sequence 811 AA;

Query Match 13.2%; Score 71; DB 23; Length 811;
Best Local Similarity 27.8%; Pred. No. 18;
Matches 22; Conservative 11; Mismatches 20; Indels 26; Gaps 2;

QY 14 CQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSGSSASASASLEWWEK----- 67
Db 560 CRIEIEKTLR-----MAELPLLTKKAFKACSQYNCAEMLEKVKPYCK 605

QY 68 -----DLEERLNSIDHDM 80
Db 606 ESLAYVTSKERLVSLEDEL 624

RESULT 11
ABB71141
ID ABB71141 standard; Protein; 1833 AA.

XX ABB71141;

AC ABB71141;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 40215.

DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL15244.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure; SEQ ID NO 40215; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1833 AA;

Query Match 13.2%; Score 71; DB 22; Length 1833;

Best Local Similarity 23.2%; Pred. No. 57;
Matches 29; Conservative 21; Mismatches 41; Indels 34; Gaps 4;

QY 12 NRCQQQAPSTLRTVTMAEFRRVPLPP-----MAEVPMLSTQNSMG----- 52

```
Db 211 DRPERKSGTSLNQAQADQKRAALPPKKVAVASTTTTASSSSSSSTSLKTSNSTSASNEV 270
QY 53 -----SSASASASLEME-----KLEER---LNSIDHDMNNKFGSGELKSMFNOG 97
Db 271 KVVTSSTSSSTSSSVRRKEADSVASKEIKROTVPAAISHSNSTSSSTASTASKSQDTNG 330
QY 98 KVEEM 102
Db 331 MQEOM 335

RESULT 12
AAG61796
ID AAG61796 standard; protein; 166 AA.
XX
AC AAG61796;
XX
DT 02-JUL-2001 (first entry)
XX
DE Interferon alpha homologue protein #7.
XX
KW IFN: interferon alpha; homologue; tumour; cancer; RNA virus;
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
KW type I diabetes; lupus erythematosus.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200125438-A2.
XX
PD 12-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27781.
XX
PR 07-OCT-1999; 99US-0415183.
XX
PA (MAXY-) MAXYGEN INC.
XX
PI Heinrichs V, Chen T, Patten PA;
DR WPI: 2001-335563/35.
XX
Novel isolated or recombinant alpha-interferon homologues useful for
PT inhibiting growth of tumors or inhibiting replication of virus and for
PT treating autoimmune diseases such as multiple sclerosis, rheumatoid
PT arthritis
XX
PS Claim 1; Page 117-118; 209pp; English.
XX
The present invention relates to interferon (IFN) alpha homologues.
CC The homologues were generated using shuffled interferon alpha
CC libraries. The homologues may be used for inhibiting
CC growth of tumour cells e.g. human carcinoma and
CC leukemia cells, inhibiting replication of RNA viruses,
CC and for treating an autoimmune disorder such as multiple
CC sclerosis, rheumatoid arthritis, type I diabetes, lupus
CC erythematosus. The present sequence is an interferon
CC alpha homologue of the invention.
XX
SQ Sequence 166 AA;
Query Match 13.1%; Score 70.5; DB 22; Length 166;
Best Local Similarity 28.0%; Pred. No. 2.1;
Matches 23; Conservative 16; Mismatches 24; Indels 19; Gaps 3;
QY 1 MONSDYFYAQ-----NRCQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSMGSSA 55
Db 30 LKDRQDFGFPQEFDCNRFQAKAISVLHEMIQOTFN-----LFSTKN---SSA 75
QY 56 SASASLEMEKDLERLNSID 77
Db 76 AWEQSLLEKFSSTELYQOINDLE 97
```

RESULT 13

```
ABB62448
ID ABB62448 standard; Protein; 176 AA.
XX
AC ABB62448;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 14136.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI: 2001-656860/75.
DR N-PSDB; ABL06551.
XX
New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
PS Disclosure; SEQ ID NO 14136; 21pp + Sequence Listing; English.
XX
The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 176 AA;
Query Match 13.0%; Score 70; DB 22; Length 176;
Best Local Similarity 27.9%; Pred. No. 2.7;
Matches 29; Conservative 15; Mismatches 48; Indels 12; Gaps 4;
QY 2 QNSQDYFYAQNRCQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTQNSMGSSASAS 61
Db 18 QQAQEQRVQVQEQQQEQQP-----TEAVPEKRIRPLTPPAEEP---GQNCNPDPAPNRL 70
QY 62 LEWWEK--DLEERLNSIDHDMNNKFGSGELKSMFNOGKVEEMD 103
Db 71 LEALQKIMELQALDAFEQDLNDR---DGYAAGAAEAEDVEESD 111

RESULT 14
ABB70398
ID ABB70398 standard; Protein; 624 AA.
XX
AC ABB70398;
XX
DT 26-MAR-2002 (first entry)
```

XX DE Drosophila melanogaster polypeptide SEQ ID NO 37986.
 XX KW Drosophila: developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR WPI: 2001-656860/75.
 XX DR N-PSDB; ABL14501.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions.
 XX PS Disclosure: SEQ ID NO 37986; 2lpp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
 XX CC useful in developmental biology and in elucidating cell signalling and
 XX CC cell-cell interactions in higher eukaryotes for the development of
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
 XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX CC sequences (ABL01840-ABL16175) and the encoded proteins
 XX CC (ABB57737-ABB72072).
 XX CC The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 624 AA;
 Query Match 13.0%; Score 70; DB 22; Length 624;
 Best Local Similarity 26.7%; Pred. No. 16;
 Matches 24; Conservative 10; Mismatches 36; Indels 20; Gaps 2;
 Qy 34 VPLPPM-----AEVPMILSTQNSMGSSA-----SASASSLEMEKDLERL 73
 Db 141 VPRPFLSRAATAAAAAAALISNNSNSAGSAGOLPHSTSSSSNNNNNNNGA 200
 Qy 74 NSIDHNNKFGSGELKSMFNOGKVEEMD 103
 Db 201 NSVNSNANNNGYHSSSSQSSMDDEEDD 230
 RESULT 15
 ID ABP35715
 XX ID ABP35715 standard; Protein; 1063 AA.
 XX AC ABP35715;
 XX DT 24-JUL-2002 (first entry)
 XX DE Fungal ZBC protein sequence #141.
 XX KW Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;
 XX KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;
 XX KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;
 XX KW angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
 XX KW fungal toxin; cell surface receptor; plant growth regulator; pigment;
 XX KW insecticide; antineoplastic.

XX OS Unidentified.
 XX PN WO200224865-A2.
 XX PD 28-MAR-2002.
 XX PF 19-SEP-2001; 2001WO-US29288.
 XX PR 19-SEP-2000; 2000US-233564P.
 XX PA (MICR-) MICROBIA INC.
 XX PI Holtzman D, Madden K, Maxon M, Sherman A;
 XX DR WPI: 2002-352005/38.
 XX DR N-PSDB; ABN79904.
 XX PT New method for improving the production of a secondary metabolite e.g.
 XX PT antineoplastic agent, ergot alkaloid from a fungus involves modulation
 XX PT of the expression of at least one zinc binuclear cluster protein gene
 XX PS Disclosure: SEQ ID 294; 49pp + sequence listing; English.
 XX CC The invention relates to improving the production of a secondary
 XX CC metabolite by a fungus. This involves modulating the expression of at
 XX CC least one ZBC (zinc binuclear cluster protein) gene in a manner to
 XX CC improve the yield of the secondary metabolite. Methods of the invention
 XX CC may be used for improving the production of the secondary metabolite e.g.
 XX CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such
 XX CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),
 XX CC a fungal toxin, a glucan synthase inhibitor, gliotoxin family of compounds,
 XX CC an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such
 XX CC as ovalicin), a modulator of cell surface receptor signalling, a plant
 XX CC growth regulator, a pigment, an insecticide, or an antineoplastic
 XX CC compound. The method results in a decrease in fermentor run-time, a
 XX CC decrease in the size of the fermentor required for the production of
 XX CC equivalent amounts of the secondary metabolite, or a decrease in the
 XX CC biomass required for the production, which translates into decreased
 XX CC waste that must be handled in downstream processing. The sequences given
 XX CC in records ABP35575-ABP35722 represent ZBC proteins.
 XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained directly from WIPO at
 XX CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 1063 AA;
 Query Match 12.8%; Score 69; DB 23; Length 1063;
 Best Local Similarity 17.2%; Pred. No. 46;
 Matches 22; Conservative 22; Mismatches 28; Indels 56; Gaps 3;
 Qy 17 QOAPSTLRTVTMAEFRRVP-----LPP 38
 Db 938 RSAPSPVRNNSLPEFAQLPSRSLVSVDMLNPDYAQTNGONNSQVQSNKPKINAAQQIPT 997
 Qy 39 MAEYPMILSTQNSMGSSASASASSLEMEKDLERLNSIDHDMNN-----KFGSELKSMF 94
 Db 998 SVQVPFMNTN-----EINNKNKNINNNNNNNNSNFSATSNLGLDFV 1043
 Qy 95 NOGKVEEM 102
 Db 1044 NNGDLEL 1051

Search completed: July 16, 2003, 12:42:29
 Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:52:52 ; Search time 23 Seconds
(without alignments)
537.001 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 104

Sequence: 1 MNSQDYFYAQRNCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 451899 seqs, 118759770 residues

Word size : 0

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	104	10	US-09-814-661A-2
2	9	8.7	9	10	Sequence 2, Appl
3	9	8.7	131	9	Sequence 27, Appl
4	9	8.7	150	9	Sequence 266, App
5	8	7.7	221	9	Sequence 33, Appl
6	8	7.7	273	9	Sequence 483, App
7	8	7.7	317	9	Sequence 26, Appl
8	8	7.7	400	10	Sequence 12, Appl
9	8	7.7	400	10	Sequence 126, App
10	8	7.7	432	9	Sequence 126, App
11	8	7.7	474	9	Sequence 10363, A
12	8	7.7	474	9	Sequence 1812, Ap
13	8	7.7	474	9	Sequence 1812, Ap
14	8	7.7	474	9	Sequence 1812, Ap
15	8	7.7	474	9	Sequence 1812, Ap
16	8	7.7	523	10	Sequence 381, App
17	8	7.7	564	9	Sequence 67, Appl
18	8	7.7	564	9	Sequence 3313, Ap
19	8	7.7	726	9	Sequence 8313, Ap
					Sequence 12539, A

```

20 7.7 931 10 US-09-906-453-2
21 6.7 10 9 US-09-572-404B-1550
22 6.7 12 9 US-09-847-101B-20
23 6.7 15 10 US-09-904-599A-1
24 6.7 83 10 US-09-867-550-1466
25 6.7 109 9 US-10-001-887-120
26 6.7 130 9 US-09-764-891-4737
27 6.7 171 10 US-09-925-301-1078
28 6.7 192 10 US-09-772-656-2
29 6.7 192 10 US-09-772-656-6
30 6.7 192 10 US-09-772-656-10
31 6.7 194 9 US-09-925-299-1005
32 6.7 194 10 US-09-925-299-1005
33 6.7 232 9 US-10-156-761-10092
34 6.7 232 9 US-10-156-761-11422
35 6.7 272 9 US-10-156-761-12370
36 6.7 274 9 US-10-156-761-13223
37 6.7 302 10 US-09-799-777-41
38 6.7 316 9 US-10-156-761-12285
39 6.7 323 9 US-10-021-811-26
40 6.7 323 10 US-09-971-798-31
41 6.7 327 10 US-09-901-884-6
42 6.7 343 10 US-09-971-798-27
43 6.7 344 9 US-10-156-761-11768
44 6.7 393 9 US-10-205-823-371
45 6.7 559 9 US-10-156-761-11383

```

ALIGNMENTS

RESULT 1

```

US-09-814-661A-2
; Sequence 2, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814,661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 104
; TYPE: PRT
; ORGANISM: S. Cerevisiae
US-09-814-661A-2

```

```

Query Match      100.0%   Score 104;   DB 10;   Length 104;
Best Local Similarity 100.0%;   Pred. No. 6.1e-92;
Matches 104;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy  1  MNSQDYFYAQRNCQQQAPSTLRVTWAEFRVPLPMAEVPMLSTONSGSSASAS 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1  MNSQDYFYAQRNCQQQAPSTLRVTWAEFRVPLPMAEVPMLSTONSGSSASAS 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy  61  SLEWKEKDLERLNSIDHDMNNKFGSGELKSMFNOGKVEEMDF 104
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  61  SLEWKEKDLERLNSIDHDMNNKFGSGELKSMFNOGKVEEMDF 104
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 2

```

US-09-814-661A-27
; Sequence 27, Application US/09814661A
; Patent No. US20020151016A1
; GENERAL INFORMATION:
; APPLICANT: Rothstein, Rodney
; APPLICANT: Zhao, Xiaolan
; TITLE OF INVENTION: A SMALL PROTEIN THAT INTERACTS WITH A RIBONUCLEOTIDE REDUCTASE
; TITLE OF INVENTION: AND USES THEREOF

```

; FILE REFERENCE: 0575/56615-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/814.661A
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Peptide
; US-09-814-661A-27

Query Match 8.7%; Score 9; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 QGKVEEMDF 104
Db 1 QGKVEEMDF 9
|||||

RESULT 3
US-09-934-455-266
; Sequence 266, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jlang, Cai-Zhong
; APPLICANT: Keddie, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaira
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934.455
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 266
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-934-455-266

Query Match 8.7%; Score 9; DB 9; Length 131;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 SSASASASS 61
Db 91 SSASASASS 99
|||||

RESULT 4
US-10-156-824A-33
; Sequence 33, Application US/10156824A
; Publication No. US20030108521A1
; GENERAL INFORMATION:
; APPLICANT: Calatrava, Manuel Rosa
; TITLE OF INVENTION: Adenovirus Protein IX, its Domain Involved in Capsid Assembly,
; TITLE OF INVENTION: Transcriptional Activity, and Nuclear Reorganization
; FILE REFERENCE: 032751-065

; CURRENT APPLICATION NUMBER: US/10/156.824A
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US 60/293,974
; PRIOR FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: mutant of Ad5 pIX connected to the 7K binding
; OTHER INFORMATION: moiety through the use of a spacer
US-10-156-824A-33

Query Match 8.7%; Score 9; DB 9; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 GSSASASAS 60
Db 133 GSSASASAS 141
|||||

RESULT 5
US-10-102-806-483
; Sequence 483, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102.806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 483
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (168)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (174)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-483

Query Match 7.7%; Score 8; DB 9; Length 221;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SASASASS 61
Db 207 SASASASS 214
|||||

RESULT 6
US-10-180-375-26
; Sequence 26, Application US/10180375
; Publication No. US20030126638A1
; GENERAL INFORMATION:
; APPLICANT: Allen, William B.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Harvell, Leslie T.

APPLICANT: Helentjaris, Timothy
APPLICANT: Li, Changjiang
APPLICANT: Lowe, Keith
APPLICANT: Oliveira, Igor Cunha
APPLICANT: Shen, Bo
APPLICANT: Tarczynski, Mitchell C.
TITLE OF INVENTION: Alteration Of Oil Traits In Plants
FILE REFERENCE: BB1458 US NA1
CURRENT APPLICATION NUMBER: US/10/180,375
CURRENT FILING DATE: 2002-06-26
NUMBER OF SEQ ID NOS: 222
SOFTWARE: Microsoft Office 97
SEQ ID NO 26
LENGTH: 273
TYPE: PRT
ORGANISM: Zea mays
US-10-180-375-26

Query Match 7.7%; Score 8; DB 9; Length 273;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
DB 9 SSASASAS 16
|||||

RESULT 7

US-09-934-455-12
Sequence 12, Application US/09934455
Publication NO. US20030121070A1
GENERAL INFORMATION:
APPLICANT: Creelman, Robert
APPLICANT: Dubell, Arnold
APPLICANT: Heard, Jacqueline
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Keddie, James
APPLICANT: Pilgrim, Marsha
APPLICANT: Ratcliffe, Oliver
APPLICANT: Reuber, Lynne
APPLICANT: Riechmann, Jose Luis
APPLICANT: Yu, Guo-Liang
APPLICANT: Pineda, Omaira
TITLE OF INVENTION: Genes for Modifying Plant Traits IV
FILE REFERENCE: MBI-0025
CURRENT APPLICATION NUMBER: US/09/934,455
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/227439
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: MBI-0022
PRIOR FILING DATE: 2001-11-16
PRIOR APPLICATION NUMBER: MBI-0023
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 516
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 317
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-934-455-12

Query Match 7.7%; Score 8; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
DB 52 SSASASAS 59
|||||

RESULT 8

US-09-073-009-126

Sequence 126, Application US/09073009
Patent No. US20010012888A1
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-073-009-126

Query Match 7.7%; Score 8; DB 10; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GSSASASA 59
DB 156 GSSASASA 163
|||||

RESULT 9

US-09-793-306-126
Sequence 126, Application US/09793306
Patent No. US20020098200A1
GENERAL INFORMATION:
APPLICANT: Campos-Neto, Antonio
APPLICANT: Skeiky, Yasir
APPLICANT: Ovendale, Pamela
APPLICANT: Jen, Shyian
APPLICANT: Lodes, Michael
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
of Tuberculosis
FILE REFERENCE: 014058-008740US
CURRENT APPLICATION NUMBER: US/09/793,306
CURRENT FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 60/185,037
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/223,828
PRIOR FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 164
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 126

; LENGTH: 400
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tb470 (Mtb-40)
; NAME/KEY: MOD_RES
; LOCATION: (1)..(400)
; OTHER INFORMATION: Xaa = any amino acid
US-09-793-306-126

Query Match 7.7%; Score 8; DB 10; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 GSSASASA 59
Db 156 GSSASASA 163
|||||

RESULT 10

US-10-156-761-10363
; Sequence 10363, Application US/10156761
; Publication No. US20030119018A1

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 245-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10363
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10363

Query Match 7.7%; Score 8; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SASASASS 61
Db 141 SASASASS 148
|||||

RESULT 11

US-09-736-457-1812
; Sequence 1812, Application US/09736457
; Patent No. US20020168637A1

; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedrick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736.457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1812

Query Match 7.7%; Score 8; DB 9; Length 474;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SASASASS 61
Db 266 SASASASS 273
|||||

RESULT 12

US-09-902-941-1812
; Sequence 1812, Application US/09902941
; Patent No. US20020172952A1

; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1812

Query Match 7.7%; Score 8; DB 9; Length 474;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SASASASS 61
Db 266 SASASASS 273
|||||

RESULT 13

US-09-849-626-1812
; Sequence 1812, Application US/09849626
; Publication No. US20020197669A1

; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626

```
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1812

Query Match
Best Local Similarity 100.0%; DB 9; Length 474;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
Db 266 SASASASS 273

RESULT 14
US-10-017-754-1812
; Sequence 1812, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1812

Query Match
Best Local Similarity 100.0%; DB 9; Length 474;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
Db 266 SASASASS 273

RESULT 15
US-10-205-823-381
; Sequence 381, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumel
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
```

```
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 381
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-381

Query Match
Best Local Similarity 100.0%; DB 9; Length 474;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
Db 266 SASASASS 273

Search completed: July 16, 2003, 13:00:09
Job time : 24 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:45:16 ; Search time 35 Seconds
(without alignments)
395.945 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 104

Sequence: 1 MNSQDYFYAQRCCQQAP.....FGSGELKSMFNGKVEEMDF 104

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	104	21	AA190470
2	9	8.7	9	21	AA190473
3	9	8.7	131	23	AAU93095
4	9	8.7	406	22	ABB60211
5	9	8.7	877	22	ABB64464
6	9	8.7	2188	22	ABB65784
7	8	7.7	76	21	AAG29825
8	8	7.7	80	21	AAG58073
9	8	7.7	101	21	AAG58072
10	8	7.7	104	21	AAG29824

11	8	7.7	116	21	AAG29823	Arabidopsis thalia
12	8	7.7	121	22	ABB64220	Drosophila melanog
13	8	7.7	156	22	ABB70392	Drosophila melanog
14	8	7.7	184	22	ABG19775	Novel human diagno
15	8	7.7	208	22	ABB64683	Drosophila melanog
16	8	7.7	221	21	AB58775	Breast and ovarian
17	8	7.7	279	21	AAG32271	Arabidopsis thalia
18	8	7.7	279	21	AAG46469	Arabidopsis thalia
19	8	7.7	285	21	AAG08088	Arabidopsis thalia
20	8	7.7	312	21	AAG32270	Arabidopsis thalia
21	8	7.7	312	21	AAG46468	Arabidopsis thalia
22	8	7.7	317	21	AAG32269	Arabidopsis thalia
23	8	7.7	317	21	AAG46467	Arabidopsis thalia
24	8	7.7	317	23	AAU92968	Arabidopsis transc
25	8	7.7	318	21	AAG08087	Arabidopsis thalia
26	8	7.7	323	21	AAG08086	Arabidopsis thalia
27	8	7.7	367	21	AAG53552	Arabidopsis thalia
28	8	7.7	393	21	AAG44569	Arabidopsis thalia
29	8	7.7	400	20	AAW73764	M. tuberculosis an
30	8	7.7	400	20	AAW73654	M. tuberculosis an
31	8	7.7	400	22	AAU08225	Mycobacterium tube
32	8	7.7	414	21	AAG44130	Arabidopsis thalia
33	8	7.7	418	21	AAG53551	Arabidopsis thalia
34	8	7.7	421	21	AAG53550	Arabidopsis thalia
35	8	7.7	444	21	AAG44568	Arabidopsis thalia
36	8	7.7	447	21	AAG44567	Arabidopsis thalia
37	8	7.7	465	21	AAG44129	Arabidopsis thalia
38	8	7.7	468	21	AAG44128	Arabidopsis thalia
39	8	7.7	474	23	AAU85550	Clone #24928 (L978
40	8	7.7	523	14	AAU40868	Flavonoid-3',5'-hy
41	8	7.7	620	22	ABB69938	Drosophila melanog
42	8	7.7	651	17	AAU95605	STR6 (suppressor o
43	8	7.7	725	21	AAG53542	Arabidopsis thalia
44	8	7.7	738	19	AAW56163	New DNA sequence i
45	8	7.7	787	22	ABB64770	Drosophila melanog

ALIGNMENTS

```
RESULT 1
AA190470
ID AA190470 standard; Protein; 104 AA.
XX
AC AA190470;
XX
DT 15-AUG-2000 (first entry)
XX
DE Yeast Sm11 (suppressor of mecl lethality) protein.
XX
KW Sm11 protein; YML058w; suppressor of mecl lethality; chromosome XIII;
KW yeast; ribonucleotide reductase inhibitor; RNR; cell division;
KW deoxynucleotide triphosphate synthesis; Mec1; Rad53;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT.
XX
OS Saccharomyces cerevisiae.
XX
PN W0200017225-A2.
XX
PD 30-MAR-2000.
XX
PF 24-SEP-1999; 99WO-US222360.
XX
PR 24-SEP-1998; 98US-0158858.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Rothstein R, Zhao X;
XX
DR WPI; 2000-283539/24.
XX
N-PSDB; AAA14263.
```

PT New Sml1 protein and its homologs, useful for treating cancer,
PT microbial infection and ataxia telangiectasia and in screening for
XX specific modulators
PS Claim 2: Fig 1C; 98pp; English.
XX
CC This sequence represents the yeast Sml1 (suppressor of mecl lethality)
CC protein. This protein is encoded by the SML1 gene, located on chromosome
CC XIII, and is an inhibitor of ribonucleotide reductase. The SML1 gene
CC (assigned the name YML058w by the Stanford yeast genomic project) was
CC initially isolated in a mecl-1 strain by genetic analysis. Sml1 is a
CC suppressor of the mecl mutant in yeast which is associated with abnormal
CC levels of recombination in both meiosis and mitosis. Sml1 also permits
CC cell growth in the absence of the cell cycle checkpoint proteins Mecl
CC and Rad53, which, unlike most checkpoint proteins, are essential for cell
CC growth. Sml1 binds to the largest subunit of ribonucleotide reductase
CC (RNR) which catalyses the rate-limiting step of deoxynucleotide
CC triphosphate (dNTP) synthesis, conversion of NTPs to dNTPs. Binding of
CC Sml1 reduces the activity of RNR and thus inhibits the synthesis of
CC dNTPs. The SML1 gene, or fragments thereof may be used to isolate the
CC human homologue of Sml1. The Sml1 protein, and compounds that modulate
CC the interaction of Sml1 with ribonucleotide reductase (RNR), may be used
CC to alter the rate at which cells divide. These are particularly useful
CC for treating cancer, microbial infection and ataxia telangiectasia (AT),
CC a condition in which patients are prone to cancer. Sml1 may also be used
CC to screen for modulatory agents, to raise specific antibodies, and for
CC stimulating the function of the ATM gene (a mammalian Mecl homologue
CC which is mutated in AT). Anti-Sml1 antibodies are used as diagnostic and
CC analytical immunassay reagents and to remove Sml1 from serum or to
CC titrate Sml1 intracellularly.
XX
SQ Sequence 104 AA;
Query Match 100.0%; Score 104; DB 21; Length 104;
Best Local Similarity 100.0%; Pred. No. 3.3e-95;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MONSDYFYAQRCCQQCAPSTLRTVTMAEFRVPLPPMAEVPMLSTONSGSSASAS 60
|||||
DB 1 MONSDYFYAQRCCQQCAPSTLRTVTMAEFRVPLPPMAEVPMLSTONSGSSASAS 60
|||||
QY 61 SLEMEKDLERLNSIDHDMNNKFGSGELKSMFNOGKVEEMDF 104
|||||
DB 61 SLEMEKDLERLNSIDHDMNNKFGSGELKSMFNOGKVEEMDF 104
|||||
RESULT 2
AAY90473
ID AAY90473 standard; Protein; 9 AA.
XX
AC AAY90473;
XX
DT 15-AUG-2000 (first entry)
XX
DE Yeast Sml1 C-terminal peptide.
XX
KW Sml1 protein; ribonucleotide reductase inhibitor; RNR; yeast;
KW cell division; deoxynucleotide triphosphate synthesis;
KW cell cycle checkpoint; cancer; microbial infection;
KW ataxia telangiectasia; AT.
XX
OS Saccharomyces cerevisiae.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl moiety"
XX
XX WO200017225-A2.
PN
XX 30-MAR-2000.
PD
XX 24-SEP-1999; 99WO-US22260.
PF

XX
XX 24-SEP-1998; 98US-0158858.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Rothstein R, Zhao X;
XX
XX WPI; 2000-283539/24.
DR
XX
XX New Sml1 protein and its homologs, useful for treating cancer,
PT microbial infection and ataxia telangiectasia and in screening for
PT specific modulators
XX
XX Example 6; Page 52; 98pp; English.
PS
XX The invention relates to the yeast SML1 (suppressor of mecl lethality)
CC gene (AAA14263) and to the Sml1 protein (AAY90470). The Sml1 gene,
CC located on chromosome XIII, encodes Sml1, an inhibitor of ribonucleotide
CC reductase. The SML1 gene (assigned the name YML058w by the Stanford yeast
CC genomic project) was initially isolated in a mecl-1 strain by genetic
CC analysis. Sml1 is a suppressor of the mecl mutant in yeast which is
CC associated with abnormal levels of recombination in both meiosis and
CC mitosis. Sml1 also permits cell growth in the absence of the cell cycle
CC checkpoint proteins Mecl and Rad53, which, unlike most checkpoint
CC proteins, are essential for cell growth. Sml1 binds to the largest
CC subunit of ribonucleotide reductase (RNR) which catalyses the
CC rate-limiting step of deoxynucleotide triphosphate (dNTP) synthesis,
CC conversion of NTPs to dNTPs. Binding of Sml1 reduces the activity of RNR
CC and thus inhibits the synthesis of dNTPs. The SML1 gene, or fragments
CC thereof may be used to isolate the human homologue of SML1. The Sml1
CC protein, and compounds that modulate the interaction of Sml1 with
CC ribonucleotide reductase (RNR), may be used to alter the rate at which
CC cells divide. These are particularly useful for treating cancer.
CC microbial infection and ataxia telangiectasia (AT), a condition in which
CC patients are prone to cancer. Sml1 may also be used to screen for
CC modulatory agents, to raise specific antibodies, and for stimulating the
CC function of the ATM gene (a mammalian Mecl homologue which is mutated in
CC AT). Anti-Sml1 antibodies are used as diagnostic and analytical
CC immunassay reagents and to remove Sml1 from serum or to titrate Sml1
CC intracellularly. Sequences AAY90471-Y90473 represent the C-terminal 9
CC amino acids of yeast ribonucleotide reductase 2 (RNR2), yeast RNR4 and
CC yeast Sml1, respectively.
XX
SQ Sequence 9 AA;
Query Match 8.7%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 96 QGKVEEMDF 104
|||||
DB 1 QGKVEEMDF 9
|||||
RESULT 3
AAU93095
ID AAU93095 standard; Protein; 131 AA.
XX
AC AAU93095;
XX
XX 02-JUL-2002 (first entry)
DT
XX Arabidopsis transcription factor #133.
DE
XX
KW Agriculture; metabolic chemical; environmental stress; drought;
KW microbial disease resistance; herbicide resistance; seed yield;
KW fruit yield; growth rate; leaf senescence; flower senescence.
KW plant; transcription factor; transgenic.
XX
OS Arabidopsis thaliana.
XX
PN WO200215675-A1.
XX

PD 28-FEB-2002.
XX
PF 22-AUG-2001; 2001WO-US26189.
XX
PR 22-AUG-2000; 2000US-227439P.
PR 16-NOV-2000; 2000US-0713994.
PR 16-APR-2001; 2001US-0837944.
XX
(MEND-) MENDEL BIOTECHNOLOGY INC.
PA (PILG/) PILGRIM M.
PA (CREE/) CREELMAN R.
PA (DUBE/) DUBELL A. J.
PA (HEAR/) HEARD J.
PA (JIAN/) JIANG C.
PA (KEDD/) KEDDIE J.
PA (ADAM/) ADAM L.
PA (RATC/) RATCLIFF O.
PA (REUB/) REUBER J. L.
PA (RIEC/) RIECHMANN J. L.
PA (YUGG/) YU G.
PA (PINE/) PINEDA O.
XX
Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX
DR WPI: 2002-292022/33.
DR N-PSDB: ABK65281.
XX
XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -
XX
PS Claim 40; Page 547-548; 94lpp; English.
XX
CC The invention relates to 1 of 232 isolated or recombinant polynucleotides
CC encoding an Arabidopsis thaliana transcription factor, their variants,
CC complements, fragments, or related polynucleotide with 31% to 95%
CC sequence identity, where the plant possesses an altered trait as compared
CC to a wild-type or reference plant, or the plant exhibits an altered
CC phenotype as compared to a wild-type or reference plant, or the plant
CC exhibits ectopic expression or altered expression of one or more genes
CC associated with a plant trait as compared to a wild plant. Also included
CC are a transgenic plant comprising the polynucleotides, a computer
CC readable medium having stored sequence information, and identifying a
CC homologue sequence from a database comprising a plurality of known plant
CC sequences comprising inputting sequence information selected from one of
CC 464 fully defined sequences given in the specification. The isolated or
CC recombinant polynucleotide is used for producing a plant having a
CC modified trait. The method comprising selecting a polynucleotide that
CC encodes a polypeptide or an antisense nucleic acid, inserting the
CC polynucleotide or antisense nucleic acid into an expression vector,
CC introducing the vector into a plant or a cell of a plant to overexpress
CC the polypeptide or antisense nucleic acid, thereby producing a modified
CC plant, and selecting for a modified trait (e.g. increased
CC production of agriculturally useful proteins or metabolic chemicals,
CC pest tolerance, environmental stress response (e.g. drought), microbial
CC disease resistance, herbicide resistance, seed and fruit yield, growth
CC rate, leaf and flower senescence and many other traits listed in the
CC specification). The present sequence is one of the 232 proteins which are
CC A. thaliana transcription factors.
XX
SQ Sequence 131 AA;

Query Match 8.7%; Score 9; DB 23; Length 131;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASASS 61
| | | | | | | |
DB 91 SSASASASS 99

RESULT 4
AB860211

ID ABB60211 standard; Protein; 406 AA.
XX
AC ABB60211;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7425.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB; ABL04314.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 7425; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 406 AA;

Query Match 8.7%; Score 9; DB 22; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASASS 61
| | | | | | | |
DB 379 SSASASASS 387

RESULT 5
AB864464
ID ABB64464 standard; Protein; 877 AA.
XX
AC ABB64464;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 20184.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 23-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 25-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 7.7%; Score 8; DB 21; Length 76;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 SASASASS 61
 |||||
 Db 51 SASASASS 58

RESULT 8

AAG58073
 ID AAG58073 standard; Protein; 80 AA.

XX AAG58073;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 74919.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.

```
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151330.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 23-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.

PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match 7.7%; Score 8; DB 21; Length 80;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SSASASAS 60
DB 47 SSASASAS 54

RESULT 9
AAG58072
ID AAG58072 standard; Protein; 101 AA.
XX AC AAG58072:
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 74918.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
```

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

```
PR 29-OCT-1999; 99US-0162142.
Query Match 7.7%; Score 8; DB 21; Length 101;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 SSASASAS 60
Db 68 SSASASAS 75
RESULT 10
AAG29824
ID AAG29824 standard; Protein; 104 AA.
XX
AC AAG29824;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35549.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
```

[illegible]

Db	79 SASASASS 86	
RESULT 11		
AAG29823		
ID	AAG29823 standard; Protein;	
XX	AC	
XX	AAG29823;	
DT	17-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein	
XX		
KW	Protein identification; sig	
KW	hybridisation assay; geneti	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439	
XX		
PR	25-FEB-1999;	99US-01218259;
PR	03-MAR-1999;	99US-01231800;
PR	03-MAR-1999;	99US-01232548;
PR	23-MAR-1999;	99US-01257988;
PR	25-MAR-1999;	99US-01262654;
PR	25-MAR-1999;	99US-01267855;
PR	01-APR-1999;	99US-01274622;
PR	01-APR-1999;	99US-01282234;
PR	08-APR-1999;	99US-01287714;
PR	16-APR-1999;	99US-01298455;
PR	19-APR-1999;	99US-01300777;
PR	21-APR-1999;	99US-01304449;
PR	23-APR-1999;	99US-01305100;
PR	23-APR-1999;	99US-01308911;
PR	28-APR-1999;	99US-01314449;
PR	30-APR-1999;	99US-01320484;
PR	30-APR-1999;	99US-01324077;
PR	04-MAY-1999;	99US-01324884;
PR	05-MAY-1999;	99US-01324885;
PR	05-MAY-1999;	99US-01324866;
PR	06-MAY-1999;	99US-01324877;
PR	07-MAY-1999;	99US-01328653;
PR	11-MAY-1999;	99US-01334256;
PR	14-MAY-1999;	99US-01342118;
PR	14-MAY-1999;	99US-01342119;
PR	14-MAY-1999;	99US-01342211;
PR	14-MAY-1999;	99US-01343707;
PR	18-MAY-1999;	99US-01347686;
PR	19-MAY-1999;	99US-01349411;
PR	20-MAY-1999;	99US-01351124;
PR	21-MAY-1999;	99US-01353553;
PR	24-MAY-1999;	99US-01356239;
PR	25-MAY-1999;	99US-01360212;
PR	27-MAY-1999;	99US-01363932;
PR	28-MAY-1999;	99US-01367822;
PR	01-JUN-1999;	99US-01372222;
PR	03-JUN-1999;	99US-01375228;
PR	04-JUN-1999;	99US-01375020;
PR	07-JUN-1999;	99US-01377244;
PR	08-JUN-1999;	99US-01380394;
PR	10-JUN-1999;	99US-01385400;
PR	10-JUN-1999;	99US-01388477;
PR	14-JUN-1999;	99US-01391119;
PR	16-JUN-1999;	99US-01394524;
PR	16-JUN-1999;	99US-01394533;
PR	17-JUN-1999;	99US-01394929;
PR	18-JUN-1999;	99US-01394454;

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140591.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.7%; Score 8; DB 21; Length 116;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
| | | | | | | |
Db 91 SASASASS 98

RESULT 12
ABB64220
ID ABB64220 standard; Protein; 121 AA.

XX ABB64220;
AC
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 19452.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
PR
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL08323.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 19452; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 121 AA;
XX
Query Match 7.7%; Score 8; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 54 SASASASS 61
Db 109 SASASASS 116
XXXXXXXXXXXX
RESULT 13
ABB70392
ID ABB70392 standard; Protein; 156 AA.
XX
AC ABB70392;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 37968.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX

XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
PR
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL14495.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 37968; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 156 AA;
XX
Query Match 7.7%; Score 8; DB 22; Length 156;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 53 SSASASAS 60
Db 56 SSASASAS 63
XXXXXXXXXXXX
RESULT 14
ABG19775
ID ABG19775 standard; Protein; 184 AA.
XX
AC ABG19775;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #19766.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
KW
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR

DR N-PSDB; AAS83962.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 50134; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 184 AA;

Query Match 7.7%; Score 8; DB 22; Length 184;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SASASASS 61
 |||||
 Db 118 SASASASS 125

RESULT 15
 ABB64683
 ID ABB64683 standard; Protein; 208 AA.
 XX
 AC ABB64683;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 20841.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL08786.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 20841; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins.
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 208 AA;
 Query Match 7.7%; Score 8; DB 22; Length 208;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 53 SSASASAS 60
 |||||
 Db 67 SSASASAS 74

Search completed: July 16, 2003, 12:51:52
 Job time : 36 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:39:36 ; Search time 40 Seconds
(without alignments)
249.949 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 537

Sequence: 1 MONSQDYFYAQNRCQQQQAP.....FGSGELKSMFNQCKVEEMDF 104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	537	100.0	104	2 S49803	hypothetical prote
2	80.5	15.0	931	2 T31731	hypothetical prote
3	80	14.9	1983	2 T00385	KIAA0624 protein
4	78.5	14.6	451	2 D84091	hypothetical prote
5	77.5	14.4	418	2 S76899	hypothetical prote
6	76.5	14.2	510	2 A45338	connexin-56 - chic
7	71.5	13.3	133	2 S69322	hypothetical prote
8	71.5	13.3	1819	2 T32008	hypothetical prote
9	71	13.2	456	2 B72130	frame-shift with c
10	71	13.2	774	2 B86492	hypothetical prote
11	71	13.2	774	2 H81540	hypothetical prote
12	71	13.2	811	2 E72003	hypothetical prote
13	70.5	13.1	541	2 AG0173	probable paraquat-
14	69	12.8	455	2 C86393	TiK7.2 protein - A
15	69	12.8	507	2 C88110	protein T24E12.1 l
16	68.5	12.8	501	2 T23011	hypothetical prote
17	68.5	12.8	785	2 S54016	SOK2 protein - yea
18	68.5	12.8	1004	2 A55142	myosin-light-chain
19	68.5	12.8	1068	2 S64015	pleiotropic drug r
20	68.5	12.8	1331	2 AE1843	hypothetical prote
21	68	12.7	330	1 ZHBPA3	gene H protein - p
22	68	12.7	332	2 JC4807	core protein H - p
23	68	12.7	486	1 A35667	Ty transcription a
24	68	12.7	1097	2 T13033	cyclin T - fruit f
25	68	12.7	1621	2 T15264	hypothetical prote
26	68	12.7	2843	1 RBHUAP	adenomatous polyo
27	67.5	12.6	330	2 T38374	probable bzip tran
28	67.5	12.6	504	2 S33194	phase-1 flagellin
29	67.5	12.6	504	2 S33186	phase-1 flagellin

30	67.5	12.6	504	2 S33190	phase-1 flagellin
31	67.5	12.6	504	2 S33193	phase-1 flagellin
32	67.5	12.6	504	2 S33191	phase-1 flagellin
33	67.5	12.6	504	2 S33189	phase-1 flagellin
34	67.5	12.6	504	2 S33188	phase-1 flagellin
35	67.5	12.6	504	2 S33187	phase-1 flagellin
36	67.5	12.6	507	2 S33192	phase-1 flagellin
37	67.5	12.6	507	2 S33185	phase-1 flagellin
38	67.5	12.6	508	2 A53465	phase-1 flagellin
39	67.5	12.6	571	2 D86164	hypothetical prote
40	67	12.5	349	2 F83623	probable methyl est
41	67	12.5	948	2 C83060	hypothetical prote
42	67	12.5	1313	2 F96673	hypothetical prote
43	67	12.5	1532	2 H96795	receptor tyrosine
44	67	12.5	2051	2 T30938	probable ABC-type
45	66.5	12.4	254	2 T35994	

ALIGNMENTS

RESULT 1

S49803

hypothetical protein YML058w - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein YM9958.04

C;Species: Saccharomyces cerevisiae

C;Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 19-Apr-2002

C;Accession: S49803

R;Devlin, K.; Churcher, C.

submitted to the EMBL Data Library, November 1994

A;Reference number: S49800

A;Accession: S49803

A;Molecule type: DNA

A;Residues: 1-104 <DEV>

A;Cross-references: EMBL:Z46729; NID:g577134; PIDN:CAA86717.1; PID:g577138; GSPDB:GNO

C;Genetics:

A;Gene: SGD:SML1; MIPS:YML058w

A;Cross-references: SGD:S0004523

A;Map position: 13L

C;Superfamily: Saccharomyces hypothetical protein YML058w

Query Match 100.0%; Score 537; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.3e-47;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MONSQDYFYAQNRCQQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTONSMGSSASAS 60

Db 1 MONSQDYFYAQNRCQQQQAPSTLRTVTMAEFRRVPLPPMAEVPMLSTONSMGSSASAS 60

Qy 61 SLEWKEKDLERLNSIDHDMNNKFGSGELKSMFNQCKVEEMDF 104

Db 61 SLEWKEKDLERLNSIDHDMNNKFGSGELKSMFNQCKVEEMDF 104

RESULT 2

T31731

hypothetical protein T05C3.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C;Accession: T31731

R;Blanchard, M.; Bradshaw, H.; Kramer, J.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of C. elegans cosmid T05C3.

A;Reference number: Z21076

A;Accession: T31731

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-931 <BLA>

A;Cross-references: EMBL:AF016428; PIDN:AAB65360.1; GSPDB:GNO0023; CESP:T05C3.3

A;Experimental source: strain Bristol N2; clone T05C3

C;Genetics:

A;Gene: CESP:T05C3.3

A;Map position: 5

1

RESULT 7
S69322
hypothetical protein YLR437c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L9753.7-a
C:Species: Saccharomyces cerevisiae
C>Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 19-Apr-2002
C:Accession: S69322
R:Du, Z.
submitted to the EMBL Data Library, February 1995
A:Description: The sequence of S. cerevisiae cosmid 9753.
A:Reference number: S3391
A:Accession: S69322
A:Molecule type: DNA
A:Residues: 1-133 <DUZ>
A:Cross-references: EMBL:U21094; NID:g665967; PIDN:AAB67525.1; PID:g2340996; GSPDB:GN0000
C:Genetics:
A:Gene: MIPS:YLR437c
A:Cross-references: SGD:S0004429
A:Map position: 12R

Query Match 13.3%; Score 71.5; DB 2; Length 133;
Best Local Similarity 35.3%; Pred. No. 3.6;
Matches 24; Conservative 7; Mismatches 26; Indels 11; Gaps 4;

QY 6 DYFYAQRRCQQQA---PSTLR---TVMAEFRVRPLPPMAEV-PMLSTQNSMGSSASAS 58
Db 51 DQGYASRTPSTSDASIQPGVIRDYSSVIVPQTRSPLTANSPLPMLINQRTM----STE 106

QY 59 ASSLEWWE 66
Db 107 ASSLEKWD 114

RESULT 8
T32008
hypothetical protein K10G6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32008
R:Davidson, S.; Wohlmann, P.; Mullen, G.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid K10G6.
A:Reference number: Z21111
A:Accession: T32008
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1819 <DAV>
A:Cross-references: EMBL:AF016659; PIDN:AAB66098.1; GSPDB:GN000020; CESP:K10G6.3
A:Experimental source: strain Bristol N2; clone K10G6
C:Genetics:
A:Gene: CESP:K10G6.3
A:Map position: 2
A:Introns: 41/2; 191/2; 556/2; 701/2; 1645/2; 1700/2; 1768/1; 1797/3

Query Match 13.3%; Score 71.5; DB 2; Length 1819;
Best Local Similarity 21.5%; Pred. No. 97;
Matches 23; Conservative 14; Mismatches 51; Indels 19; Gaps 2;

QY 10 AQRRCQQQAAPSTLTVTVMAEFRVRPLPPMAEV-PMLSTQNSMGSSASASASLEWKEKDL 69
Db 985 APSRQQQQOOPP-----VAYQVQFNGRPLPPMQLPPLQNPHNHQOQQHQLHQSOMNYQQVQQ 1040

QY 70 ERLNSID-----HDMNNKFGSGELKSMFNQKVEE 101
Db 1041 VQOVQHVQOQQNLQNHQHHQOQQHQQOQQOQQAPGNRSRSHSNVCKWEQ 1087

RESULT 9
B72130
frame-shift with cpn0010 - Chlamydomophila pneumoniae (strain CWL029)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: B72130
R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
N:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: B72130
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-456 <ARN>
A:Cross-references: GB:AE001586; GB:AE001363; NID:g4376263; PIDN:AAD18169.1; PID:g437
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: CPn0010.1

Query Match 13.2%; Score 71; DB 2; Length 456;
Best Local Similarity 27.8%; Pred. No. 19;
Matches 22; Conservative 11; Mismatches 20; Indels 26; Gaps 2;

QY 14 CQOQQAAPSTLTVTVMAEFRVRPLPPMAEV-PMLSTQNSMGSSASASASLEWKEK-----67
Db 208 CRIEIEKTLR-----MAELPLPTTKAFKACSOYNCAEMLEKVKPYCK 253

QY 68 -----DLEERLNSIDHDM 80
Db 254 ESLAVVTSKERLVSLDEDL 272

RESULT 10
B86492
hypothetical protein CPJ0010 [imported] - Chlamydomophila pneumoniae (strain J138)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: B86492
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: B86492
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <STO>
A:Cross-references: GB:BA000008; NID:g8978383; PIDN:BA098220.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0010

Query Match 13.2%; Score 71; DB 2; Length 774;
Best Local Similarity 27.8%; Pred. No. 37;
Matches 22; Conservative 11; Mismatches 20; Indels 26; Gaps 2;

QY 14 CQOQQAAPSTLTVTVMAEFRVRPLPPMAEV-PMLSTQNSMGSSASASASLEWKEK-----67
Db 523 CRIEIEKTLR-----MAELPLPTTKAFKACSOYNCAEMLEKVKPYCK 568

QY 68 -----DLEERLNSIDHDM 80
Db 569 ESLAVVTSKERLVSLDEDL 587

RESULT 11
H81540
hypothetical protein CP0764 [imported] - Chlamydomophila pneumoniae (strain AR39)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: H81540
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: H81540
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-774 <REA>
A:Cross-references: GB:AE002236; GB:AE002161; NID:g7189675; PIDN:AAF38564.1; PID:g718967
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0764

Query Match 13.2%; Score 71; DB 2; Length 774;
Best Local Similarity 27.8%; Pred. No. 37;
Matches 22; Conservative 11; Mismatches 20; Indels 26; Gaps 2;
QY 14 COOQAPSTLRVTVMAEFRVPLPPMAEVPMLSTQNSMGSSASASASLEMMWEK----- 67
I::: |||
|||::|::: | | || ||
Db 523 CRIBIEKTLR-----MAELPLLPTRKKAFAKACSQNSCAEMLEKVKPYCK 568

QY 68 -----DLEERLNSIDHDM 80
: ||| |::|
Db 569 ESLAYVTSKERLVSLDEDL 587

RESULT 12
E72003
Hypothetical protein - Chlamydomydia pneumoniae (strain CWL029)
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: E72003
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: E72003
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-811 <ARN>
A:Cross-references: GB:AE001685; GB:AE001363; NID:g4377378; PIDN:AAD19191.1; PID:g437738
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: Cpn1054

Query Match 13.2%; Score 71; DB 2; Length 811;
Best Local Similarity 27.8%; Pred. No. 39;
Matches 22; Conservative 11; Mismatches 20; Indels 26; Gaps 2;
QY 14 COOQAPSTLRVTVMAEFRVPLPPMAEVPMLSTQNSMGSSASASASLEMMWEK----- 67
I::: |||
|||::|::: | | || ||
Db 560 CRIBIEKTLR-----MAELPLLPTRKKAFAKACSQNSCAEMLEKVKPYCK 605
QY 68 -----DLEERLNSIDHDM 80
: ||| |::|
Db 606 ESLAYVTSKERLVSLDEDL 624

RESULT 13
AG0173
probable paraquat-inducible protein B pqlB [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AG0173
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0173
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-541 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90250.1; PID:gi5979470; GSPDB:GN00175
C:Genetics:
A:Gene: pqlB
C:Superfamily: pqlB protein

Query Match 13.1%; Score 70.5; DB 2; Length 541;
Best Local Similarity 25.0%; Pred. No. 27;
Matches 22; Conservative 18; Mismatches 35; Indels 13; Gaps 3;
QY 27 TMAEFRVPLPPM-----AEVPMSTQNSMGSSASASASLEMMWEKDLERLNS 75
I::: |||
|||::|::: | | || ||
Db 419 TLDKINNPLPNMNVETKALAESOKTMRQVKLASLNTITSSQMOELPKDLQKLTNE 478
QY 76 IDHDNNKFGSGELKSMFNQKVEEMD 103
::: |::|
Db 479 LNRSKMGFGPGSPAYNKMV--GDMQRLLD 504

RESULT 14
C86393
TIK7.2 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: C86393
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86393
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <STO>
A:Cross-references: GB:AE005172; NID:g9797768; PIDN:AAF98586.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 12.8%; Score 69; DB 2; Length 455;
Best Local Similarity 24.3%; Pred. No. 30;
Matches 28; Conservative 19; Mismatches 40; Indels 28; Gaps 4;
QY 10 AONRCQOQAPSTLRVTVMAEFRVPLPPMAEVPMLSTQN---SMGSSASASASLEMMWE 66
I::: |::|
|||::|::: | | || ||
Db 125 SQSDTEASSSPTRRS-----KRVVVKPHHKAFAVGVGNGIMNQISASSDASIEPEQ 178
QY 67 KDLEERL-----NSIDHDMNN-----KFGSGELKSMFNQKVEEM 102
::: |::|
Db 179 EDMAESLMMLSRDSSFKKGHNSLAESDNNVILETKSSSGQVKMFNVKNVEEL 233

RESULT 15
C88110
protein T24E12.1 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C88110
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C88110
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-507 <STO>
A:Cross-references: GB:chr_II; PIDN:AB95039.1; PID:g2746886; GSPDB:GN00020; CESP:T24E
C:Genetics:
A:Gene: T24E12.1
A:Map position: 2

Query Match 12.8%; Score 69; DB 2; Length 507;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 12:26:21 ; Search time 22 Seconds
(without alignments)
196.070 Million cell updates/sec

Title: US-09-814-661A-2

Perfect score: 537

Sequence: 1 MNSQDYFYAQNRCQQQAP.....FGSGELKSMFNOGKVEEMDF 104

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	537	100.0	104	1	YMF8_YEAST
2	76.5	14.2	510	1	CX56_CHICK
3	69	12.8	1063	1	PDR1_YEAST
4	68.5	12.8	785	1	SOK2_YEAST
5	68	12.7	330	1	VGH_BPAL3
6	68	12.7	332	1	VGH_BPPHK
7	68	12.7	486	1	TEC1_YEAST
8	68	12.7	1057	1	CCT_DROME
9	68	12.7	2843	1	APC_HUMAN
10	67.5	12.6	330	1	YDC3_SCHPO
11	67.5	12.6	380	1	TGT_CLOPE
12	67.5	12.6	504	1	FLIC_SALBU
13	67.5	12.6	504	1	FLIC_SALDE
14	67.5	12.6	504	1	FLIC_SALDU
15	67.5	12.6	504	1	FLIC_SALEN
16	67.5	12.6	504	1	FLIC_SALMC
17	67.5	12.6	504	1	FLIC_SALMO
18	67.5	12.6	504	1	FLIC_SALNA
19	67.5	12.6	504	1	FLIC_SALRO
20	67.5	12.6	504	1	FLIC_SALSO
21	67.5	12.6	507	1	FLIC_SALSE
22	67.5	12.6	507	1	FLIC_SALON
23	67.5	12.6	610	1	YQDA_CAEEL
24	67	12.5	351	1	APLA_HUMAN
25	67	12.5	596	1	HMEN_ANOGA
26	66.5	12.4	406	1	MYC_BRARE
27	66.5	12.4	552	1	HMD1_DICDI
28	66.5	12.4	706	1	DREB_MOUSE
29	66.5	12.4	741	1	ERF2_PICPI
30	66	12.3	284	1	FLAL_VIBPA
31	66	12.3	466	1	YL37_CAEEL
32	65.5	12.2	771	1	YJCO_YEAST
33	65	12.1	169	1	CX41_MOUSE

ALIGNMENTS

RESULT 1

YMF8_YEAST
ID YMF8_YEAST STANDARD; PRT: 104 AA.
AC Q04964;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 11.8 kDa protein in OG1-CNA2 intergenic region.
GN YML058W OR YM958.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=9288C / AB972;
RA Davlin K., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46729; CAA86717.1; -
CC SGD; S0004523; YML058W.
KW Hypothetical protein.
SQ SEQUENCE 104 AA; 11834 MW; 02735C9178E431B2 CRC64;
Query Match 100.0%; Score 537; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 3e-47;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSQDYFYAQNRCQQQAPSTLRTVTMAEFRVPLPPMAEVPMLSTQNSMGSSASAS 60
|||||
Db 1 MNSQDYFYAQNRCQQQAPSTLRTVTMAEFRVPLPPMAEVPMLSTQNSMGSSASAS 60
|||||

QY 61 SLEWKEKDLERLNSIDHDMNNKFGSGELKSMFNOGKVEEMDF 104
|||||
Db 61 SLEWKEKDLERLNSIDHDMNNKFGSGELKSMFNOGKVEEMDF 104
|||||

RESULT 2

CX56_CHICK
ID CX56_CHICK STANDARD; PRT: 510 AA.
AC P29415;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gap junction Cx56 protein (Connexin 56).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

34 65 12.1 577 1 VGLP_PVRRI
35 65 12.1 721 1 NICA_CAEEL
36 65 12.1 887 1 PMC2_MOUSE
37 65 12.1 1687 1 VIT2_FUNHE
38 64 11.9 438 1 PMCI_MOUSE
39 64 11.9 616 1 NXFI_COTJA
40 64 11.9 643 1 SGT1_ARATH
41 64 11.9 786 1 SYQ_CAEEL
42 64 11.9 1742 1 MY5C_HUMAN
43 63.5 11.8 265 1 Y280_MYCGE
44 63.5 11.8 422 1 MYC_AVIMC
45 63.5 11.8 422 1 MYC_AVIMD

P08354 pseudorabie
Q23316 caenorhabdi
P56960 mus musculu
Q98893 fundulus he
Q9jhi7 mus musculu
P58797 coturnix co
Q91sm5 arabidopsi
O62431 caenorhabdi
Q9nqx4 homo sapien
P47522 mycoplasma
P01110 avian myelo
P06295 avian myelo

Db 998 SVQVPMFTN-----EINNHNKNININNNNSNNSFNSATSFNLGTLDFV 1043

QY 95 NOKVVEEM 102
| | | | |
Db 1044 NMGDLLEL 1051

RESULT 4

ID_SOK2_YEAST STANDARD; PRT; 785 AA.

AC P53438; 21-JUL-1986 (Rel. 01, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SOK2 protein.

GN SOK2 OR YMR016C OR YMR711.03C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96069395; PubMed=8524252;

RA Ward M.P., Gimeno C.J., Fink G.R., Garrett S.;

RT "SOK2 may regulate cyclic AMP-dependent protein kinase-stimulated

RT growth and pseudohyphal development by repressing transcription.";

RL Mol. Cell. Biol. 15:6854-6863(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288c / AB972;

RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: PLAYS A GENERAL REGULATORY ROLE IN THE CYCLIC AMP-

CC DEPENDENT PROTEIN KINASE-STIMULATED (PKA) SIGNAL TRANSDUCTION

CC PATHWAY BY REGULATING THE EXPRESSION OF GENES IMPORTANT IN GROWTH

CC AND DEVELOPMENT. MAY INHIBIT THE SWITCH FROM UNICELLULAR TO

CC FILAMENTOUS GROWTH.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- SIMILARITY: BELONGS TO THE EFG1/PHD1/STUA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; S80522; AAB35749.1; -

DR EMBL; Z49211; CAA89117.1; -

DR TRANSFAC; T03616; -

DR SGD; S0004618; SOK2.

DR InterPro; IPR003163; Yeast_DNA_bind.

DR Pfam; PF02292; Yeast_DNA_bind; 1.

DR Transcription regulation; Nuclear protein; DNA-binding.

FW DOMAIN 38 45 POLY-GLN.

FT DOMAIN 148 151 POLY-TYR.

FT DOMAIN 169 172 POLY-TYR.

FT DOMAIN 206 217 POLY-GLN.

FT DOMAIN 331 335 POLY-GLN.

FT DOMAIN 526 532 POLY-SER.

SQ SEQUENCE 785 AA; 85643 MW; F4E0F224BE84645F CRC64;

Query Match 12.8%; Score 68.5; DB 1; Length 785;

Best Local Similarity 24.7%; Pred. No. 29;

Matches 24; Conservative 19; Mismatches 31; Indels 23; Gaps 5;

QY 5 QDYFYAQNRC-COOQAPSTLRTVT-MAEFRRVPLPPMAEYPMILSTQNSMGSSASASL 62

| : : | | | | | : : | : : | : : | : : | : : | : : | : : | : : |

Db 355 QETYLTAIRHGVSQQYDSQVSMKNTNSFQITIRP-----MPLATTNATGNTSGTSASI 410

| : : | | | | | : : | : : | : : | : : | : : | : : | : : | : : |

QY 63 -----EMWEKD-----LEERLNSI-----DHDMMN 82

| | | : : | | | | | | | | | | | | | | | | | | | | | | |

Db 411 IRPRVTTTWEDEKTLCYQVEANGISVVRADNDMVN 447

RESULT 5

VGH_BPAL3

ID VGH_BPAL3 STANDARD; PRT; 330 AA.

AC P03650;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Minor spike protein (H protein) (Pilot protein).

GN H.

OS Bacteriophage alpha-3.

OC Viruses; ssDNA viruses; Microviridae; Microvirus.

OX NCBI_TaxID=10849;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92223109; PubMed=1532908;

RA Kodaira K.-I., Nakano K., Okada S., Taketo A.;

RT "Nucleotide sequence of the genome of the bacteriophage alpha 3:

RT interrelationships of the genome structure and the gene products with

RT those of the phages, phi X174, G4 and phi K.;"

RL Biochim. Biophys. Acta 1130:277-288(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=85226468; PubMed=2988629;

RA Kodaira K.-I., Nakano K., Taketo A.;

RT "Function and structure of microvirid phage alpha 3 genome. DNA

RT sequence of H gene and properties of missense H mutant.";

RL Biochim. Biophys. Acta 825:255-260(1985).

RN [3]

RP SEQUENCE OF 1-70 FROM N.A.

RX MEDLINE=80049950; PubMed=387790;

RA Sims J., Capon D., Dressler D.;

RT "dnaG (primase)-dependent origins of DNA replication. Nucleotide

RT sequences of the negative strand initiation sites of bacteriophages

RT St-1, phi K, and alpha 3.;"

RL J. Biol. Chem. 254:12615-12628(1979).

CC -!- FUNCTION: MINOR SPIKE COMPONENT OF THE VIRAL SHELL. H PROTEIN IS

CC INVOLVED IN THE EJECTION OF THE PHAGE DNA IN THE HOST AND IS

CC INJECTED WITH THE DNA IN THE PERIPLASMIC SPACE OF THE HOST.

CC -!- SUBUNIT: THE VIRION IS COMPOSED OF 60 COPIES EACH OF THE F, G, AND

CC J PROTEINS, AND 12 COPIES OF THE H PROTEIN. THERE ARE 12 SPIKES

CC WHICH ARE EACH COMPOSED OF 5 G AND ONE H PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X60322; CAA42883.1; -

DR EMBL; M25640; AAA32174.1; -

DR EMBL; J02444; AAA32176.1; -

DR PIR; A04257; ZHBP33.

DR PIR; A21537; A21537.

DR PIR; S22332; S22332.

KW Coat protein.

SQ SEQUENCE 330 AA; 34844 MW; B194EF44A83CE8 CRC64;

Query Match 12.7%; Score 68; DB 1; Length 330;

Best Local Similarity 23.1%; Pred. No. 11;

Matches 21; Conservative 18; Mismatches 50; Indels 2; Gaps 2;

QY 2 QNSQDYFYAQNRC-COOQAPSTLRTVT-MAEFRRVPLPPMAEYPMILSTQNSMGSSASASA 59

| : : | | | | | : : | : : | : : | : : | : : | : : | : : | : : |

Db 196 QNTKDSVYAQNEMLQYNQESQARVASILANTDLTTKQATHEIMRMALTRAQETGQHLTN 255

| : : | | | | | : : | : : | : : | : : | : : | : : | : : | : : |

QY 60 SLEMEKLEERLNSIDHDMNNKFGSGEL 90

| : : | | | : : | | | | | | | | | | | | | | | | | |

Db 256 SQIMALEKKVYAEIGIKHQDTQNSRYGSSQV 286

[illegible]

RA VARIANTS FAP.
RX MEDLINE-93244793; PubMed-1338691;
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
RA Nakamura Y., Horii A.;
RT "Somatic mutation of the APC gene in gastric cancer: frequent
RT mutations in very well differentiated adenocarcinoma and signet-ring
RT cell carcinoma.";
RL Hum. Mol. Genet. 1:559-563(1992).
RN [11]
RP VARIANT FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 AND S-2502.
RX MEDLINE-93250848; PubMed-1338764;
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
RA Baba S., Nakamura Y.;
RT "Screening for germ-line mutations in familial adenomatous polyposis
RT patients: 61 new patients and a summary of 150 unrelated patients.";
RL Hum. Mutat. 1:467-473(1992).
RN [12]
RP VARIANT FAP TRP-99.
RC TISSUE-Peripheral blood lymphocytes;
RX MEDLINE-95134544; PubMed-7833149;
RA Dobbie Z., Spycher M., Huerliman R., Ammann T., Roth J.,
RA Mueller A., Mueller H., Scott R.J.;
RT "Mutational analysis of the first 14 exons of the adenomatous
RT polyposis coli (APC) gene.";
RL Eur. J. Cancer 30A:1709-1713(1994).
RN [13]
RP VARIANT FAP GLY-722.
RX MEDLINE-95134330; PubMed-7833931;
RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romo L., Pilia S., Prete F., Mareni C., Guanti G.;
RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
RT FAP patients.";
RL Hum. Mol. Genet. 3:1687-1698(1994).
RN [14]
RP VARIANT FAP ILE-171.
RX MEDLINE-9714176; PubMed-8990002;
RA van der Luijt R.B., Meera Khan P., Vasen H.F.A., Tops C.M.J.,
RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
RA Plug R.J., Griffioen G., Fodde R.;
RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
RT familial adenomatous polyposis: 67 germline mutations identified by
RT DGSB, PTT, and southern analysis.";
RL Hum. Mutat. 9:7-16(1997).
RN [15]
RP VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.
RX MEDLINE-98080146; PubMed-9419979;
RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
RA Muraoka M., Nagato M., Chong J.-M., Kolke M., Terada T., Kawahara Y.,
RA Fukutome A., Tomiyama J., Chuganji Y., Momoi M., Utsunomiya J.;
RT "Drastic genetic instability of tumors and normal tissues in Turcot
RT syndrome.";
RL Oncogene 15:2877-2881(1997).
RN [17]
RP VARIANT LYS-1307.
RX MEDLINE-98400248; PubMed-9731522;
RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
RA Antin-Ozerkis D., Andrusis I., Daly M., Pinsky L., Schrag D.,
RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
RT "The APC I1307K allele and breast cancer risk.";
RL Nat. Genet. 20:13-14(1998).
RN [18]
RP VARIANTS LYS-1307 AND GLN-1317.
RC TISSUE-Peripheral blood;
RX MEDLINE-98393712; PubMed-9724771;

RA Frayling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
RT "The APC variants I1307K and E1317Q are associated with colorectal
RT tumors, but not always with a family history.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
RN [19]
RP VARIANT LYS-1307.
RX MEDLINE-98400259; PubMed-9731533;
RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
RT "The APC I1307K allele and cancer risk in a community-based study of
RT Ashkenazi Jews.";
RL Nat. Genet. 20:62-65(1998).
RN [20]
RP VARIANT LYS-1307.
RX MEDLINE-99138651; PubMed-9973276;
RA Gryfe R., Di Nicola N., Lal G., Gallinger S., Redston M.;
RT "Inherited colorectal polyposis and cancer risk of the APC I1307K
RT polymorphism.";
RL Am. J. Hum. Genet. 64:378-384(1999).
RN [21]
RP VARIANTS FAP CYS-1171 AND THR-2738, AND VARIANTS GLY-1057 AND
RP VAL-1822.
RX MEDLINE-99133859; PubMed-9950360;
RA Wallis Y.L., Morton D.G., McKeown C.M., Macdonald F.;
RT "Molecular analysis of the APC gene in 205 families: extended
RT genotype-phenotype correlations in FAP and evidence for the role of
RT APC amino acid changes in colorectal cancer predisposition.";
RL J. Med. Genet. 36:14-20(1999).
RN [22]
RP VARIANT FAP PRO-1184.
RX MEDLINE-99401091; PubMed-10470088;
RA Lamum H., Ilyas M., Rowan A., Clark S., Johnson V., Bell J.A.,
RA Frayling I.M., Erstathou J., Pack K., Payne S., Roylance R.,
RA Gorman P., Sheer D., Neale K., Phillips R., Talbot I.C., Bodmer W.F.,
RA Tomlinson I.P.M.;
RT "The type of somatic mutation at APC in familial adenomatous polyposis
RT is determined by the site of the germline mutation: a new facet to
Query Match 12.7%; Score 68; DB 1; Length 2843;
Best Local Similarity 27.8%; Pred. No. 1.5e+02;
Matches 27; Conservative 17; Mismatches 33; Indels 20; Gaps 3;
Oy 18 QAPTTLRTVTWAEFRFRRVPLPPMAEVLSTONGSSA---SASASSLMEKDEERLN 74
Db 2534 ESPRLPLNRSGTWKREHSHKSSSLPRVSTWRTGSSSSILSASSESEKAKSEDEKHVN 2593
Oy 75 SIDHDMNNKFGSGELKSMFNQ-----GKVEEMDF 104
Db 2594 SI-----SCTKQSKENQVSAKGTWRKIKENEF 2620
RESULT 10
YDC3_SCHPO
ID YDC3_SCHPO STANDARD; PRT; 330 AA.
AC Q10424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C25G10.03 in chromosome 1.
GN SPAC25G10.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE-21848401; PubMed-11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford O., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B., Weltegens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs N., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffart A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas R., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RL "The genome sequence of Schizosaccharomyces pombe.";
 CC -!- SURCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z70691; CAA94632.1; -
 DR HSSP: P03069; ZDGC.
 DR InterPro: IPR004827; TF_BZIP.
 DR SMART: SM00338; BRL2; 1.
 DR PROSITE: PS00036; BZIP_BASIC; 1.
 KW Hypothetical protein; Transcription regulation; DNA-binding;
 KW Nuclear protein.
 FT DNA_BIND 270 288 BASIC MOTIF (BY SIMILARITY).
 SQ SEQUENCE 330 AA; 36189 MW; A681434C779DF960 CRC64;

Query Match 12.6%; Score 67.5; DB 1; Length 330;
 Best Local Similarity 27.0%; Pred. No. 13;
 Matches 24; Conservative 17; Mismatches 35; Indels 13; Gaps 3;

QY 11 QNRCCQQAP---STLRTVTMAEFRVPLPPMAEVPMLSTQNSMGSSASASSLEMMWEK 67
 Db 71 QEKVQQQNPKEISTLQQVKEEVSNTFSAPL-----NATGNFSSANPASIDLAVL 121
 QY 68 DLEELNSIDHDNNKFGSGELKSMFNQ 96
 Db 122 DLQKLLTLPDSKETOEXTSSQ-RELFEQ 149

RESULT 11

TGT_CLOPE
 ID TGT_CLOPE STANDARD; PRT; 380 AA.
 AC Q8XJ16;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Queuine tRNA-ribosyltransferase (EC 2.4.2.29) (tRNA-guanine
 DE transglycosylase) (Guanine insertion enzyme).
 GN Tcr OK CP1945.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium
 OX NCBI_TaxID=1502;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=13 / Type A;
 RX PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 CC -!- FUNCTION: Exchanges the guanine residue with 7-aminomethyl-7-
 CC deazaquinine in tRNAs with GU(N) anticodons (tRNA-Asp, -Asn, -His
 CC and -Tyr). After this exchange, a cyclopentendiol moiety is
 CC attached to the 7-aminomethyl group of 7-deazaquinine, resulting
 CC in the hypermodified nucleoside queuine (Q) (7-((4,5-cis-
 CC dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaquanosine (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: tRNA guanine + queuine = tRNA queuine +
 CC guanine.
 CC -!- COFACTOR: Binds and requires zinc for activity. Also requires
 CC magnesium (by similarity).
 CC -!- SIMILARITY: BELONGS TO THE QUEUINE TRNA-RIBOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AP003192; BAB81651.1; -
 DR InterPro: IPR004803; Q-trNA_tgt.
 DR InterPro: IPR002616; Que_trNAtransf.
 DR InterPro: IPR002363; Ribosomal_L10eub.
 DR TIGRFAMS: TIGR00430; Q-trNA_tgt; 1.
 DR TIGRFAMS: TIGR00449; tgt_general; 1.
 KW Queuine biosynthesis; Transferase; Glycosyltransferase;
 KW tRNA processing; Zinc; Magnesium; Complete proteome.
 FT ACT_SITE 94 94 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT METAL 313 313 ZINC (BY SIMILARITY).
 FT METAL 315 315 ZINC (BY SIMILARITY).
 FT METAL 318 318 ZINC (BY SIMILARITY).
 FT METAL 344 344 ZINC (BY SIMILARITY).
 SQ SEQUENCE 380 AA; 43200 MW; B0B7C41D254BED8 CRC64;

Query Match 12.6%; Score 67.5; DB 1; Length 380;
 Best Local Similarity 29.5%; Pred. No. 15;
 Matches 26; Conservative 14; Mismatches 33; Indels 15; Gaps 4;

QY 15 QQQQAPSTL-RVTMAEFRVPLPPMAEVPMLSTQNSMGSSASASSLEMMWEKDLERL 73
 Db 130 ESMQIQSLNGSTIAMAFDECIPNP-----STREYVEKSVARTRWLERCKEM-DRL 180
 QY 74 NSIDHDNNKFGSGELKSMFNQKVEE 101
 Db 181 NSLDDTVNKEQMLFG-----INOGGYVE 203

RESULT 12
 FLIC_SALBU
 ID FLIC_SALBU STANDARD; PRT; 504 AA.
 AC Q06969;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Flagellin (Phase-1-C flagellin).
 GN FLIC.
 OS Salmonella budapest.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=28143;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE-93374829; PubMed-76900024;
RA Masten B.J., Joys T.M.;
RT "Molecular analyses of the Salmonella g. . . flagellar antigen
RT complex.";
CC J. Bacteriol. 175:5359-5365(1993).
CC
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z15065; CAA78774.1;
CC InterPro: IPR001492; Flagellin_C.
CC Pfam: PF00669; Flagellin_N; 1.
CC PRINTS: PR00207; FLAGELLIN.
CC ProDom: PD000316; Flagellin_C; 1.
CC Flagella.
CC KW Flagella.
CC FT INIT_MET 0 0 BY SIMILARITY.
CC SQ SEQUENCE 504 AA; 52790 MW; 1FD1498751B6475E CRC64;

Query Match 12.6%; Score 67.5; DB 1; Length 504;
Best Local Similarity 29.0%; Pred. No. 21;
Matches 18; Conservative 15; Mismatches 26; Indels 3; Gaps 1;

QY 39 MAEVPMLSTQNSMGSSASASSLEWKEKLEERLNSIDHDMNNKFGSGELKSMFNQCK 98
: | | | | | : | | | | | : | | | | | : | | | | |
Dd 88 LQVRVLSVQATNGTN---SDSLKSIQDEIQORLEEDRVSNQTFNGVKVLSQDNQMK 144

QY 99 VE 100
Dd 145 IQ 146

RESULT 13
FLIC_SALDE STANDARD; PRT; 504 AA.
AC Q06370;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Flagellin (Phase-1-C flagellin).
GN FLIC.
OS Salmonella derby.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=28144;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5241;
RX MEDLINE-941195780; PubMed-8146152;
RA Li J., Nelson K., McWhorter A.C., Whittam T.S., Selander R.K.;
RT "Recombinational basis of serovar diversity in Salmonella enterica.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:2552-2556(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 6960;
RX MEDLINE-93374829; PubMed-76900024;
RA Masten B.J., Joys T.M.;
RT "Molecular analyses of the Salmonella g. . . flagellar antigen
RT complex.";
CC J. Bacteriol. 175:5359-5365(1993).
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO

```

```

CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U06225; AAA17866.1;
CC EMBL: Z15066; CAA78775.1;
CC InterPro: IPR001492; Flagellin_N.
CC Pfam: PF00669; Flagellin_N; 1.
CC PRINTS: PR00207; FLAGELLIN.
CC ProDom: PD000316; Flagellin_C; 1.
CC Flagella.
CC KW Flagella.
CC FT INIT_MET 0 0 BY SIMILARITY.
CC SQ SEQUENCE 504 AA; 52818 MW; E428EB7FF8DA790 CRC64;

Query Match 12.6%; Score 67.5; DB 1; Length 504;
Best Local Similarity 29.0%; Pred. No. 21;
Matches 18; Conservative 15; Mismatches 26; Indels 3; Gaps 1;

QY 39 MAEVPMLSTQNSMGSSASASSLEWKEKLEERLNSIDHDMNNKFGSGELKSMFNQCK 98
: | | | | | : | | | | | : | | | | | : | | | | |
Dd 88 LQVRVLSVQATNGTN---SDSLKSIQDEIQORLEEDRVSNQTFNGVKVLSQDNQMK 144

QY 99 VE 100
Dd 145 IQ 146

RESULT 14
FLIC_SALDU STANDARD; PRT; 504 AA.
AC Q06971;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Flagellin (Phase-1-C flagellin).
GN FLIC OR FLIC1.
OS Salmonella dublin.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=98360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92276338; PubMed=1592813;
RA Selander R.K., Smith N.H., Li J., Beltran P., Ferris K.E.,
RA Kopecko D.J., Rubin F.A.;
RT "Molecular evolutionary genetics of the cattle-adapted serovar
RT Salmonella dublin.";
RL J. Bacteriol. 174:3587-3592(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15480;
RX MEDLINE-93374829; PubMed=76900024;
RA Masten B.J., Joys T.M.;
RT "Molecular analyses of the Salmonella g. . . flagellar antigen
RT complex.";
CC J. Bacteriol. 175:5359-5365(1993).
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.

```


CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84973; AAA27081.1; -
CC EMBL: Z15068; CAA78776.1; -
CC InterPro: IPR001492; FlagellinN.
CC InterPro: IPR001029; Flagellin_C.
CC Pfam: PF00669; Flagellin_N; 1.
CC Pfam: PF00700; Flagellin_C; 1.
CC PRINTS: PR00207; Flagellin_C; 1.
CC ProDom: PD000316; Flagellin_C; 1.
CC Flagella.
CC KW Flagella.
CC FT INIT_MET 0 0 BY SIMILARITY.
CC SQ SEQUENCE 504 AA; 52860 MW; 34AE0BDB5033DAFC CRC64;
CC -----
CC Query Match 12.6%; Score 67.5; DB 1; Length 504;
CC Best Local Similarity 29.0%; Pred. No. 21;
CC Matches 18; Conservative 15; Mismatches 26; Indels 3; Gaps 1;
CC
CC QY 39 MAEVPMLSTQNSMGSSASASASLEMKWEDLERLNSIDHDMNNKFGSGELKSMFNQK 98
CC : | | | | | : | | | : : : : | | | : | : | | |
CC Db 88 LQVRRELSVQATNGTN---SDSLKSTQDEIQORLEEDRVSNQTFNGVKVLSQDNQMK 144
CC
CC QY 99 VE 100
CC : :
CC Db 145 IQ 146
CC
CC RESULT 15
CC FLIC_SALEN STANDARD; PRT; 504 AA.
CC AC Q06972;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE Flagellin (Phase-1-C flagellin).
CC GN FLIC.
CC OS Salmonella enteritidis.
CC SC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC OC Salmonella.
CC OX NCBI_TaxID=592;
CC [1]
CC RN SEQUENCE FROM N.A.
CC RC STRAIN=ATCC 13076;
CC RX MEDLINE=93374829; PubMed=7690024;
CC RA Masten B.J., Joys T.M.;
CC RT "Molecular analyses of the Salmonella g. . . flagellar antigen
CC complex.";
CC RL J. Bacteriol. 175:5359-5365(1993).
CC CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC CC -1- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z15068; CAA78777.1; -
CC InterPro: IPR001492; FlagellinN.

DR InterPro: IPR001029; Flagellin_C.
DR Pfam: PF00669; Flagellin_N; 1.
DR Pfam: PF00700; Flagellin_C; 1.
DR PRINTS: PR00207; FLAGELLIN.
DR ProDom: PD000316; Flagellin_C; 1.
KW Flagella.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 504 AA; 52850 MW; 4040770EEE2E9985 CRC64;

Query Match 12.6%; Score 67.5; DB 1; Length 504;
Best Local Similarity 29.0%; Pred. No. 21;
Matches 18; Conservative 15; Mismatches 26; Indels 3; Gaps 1;
QY 39 MAEVPMLSTQNSMGSSASASASLEMKWEDLERLNSIDHDMNNKFGSGELKSMFNQK 98
: | | | | | : | | | : : : : | | | : | : | | |
Db 88 LQVRRELSVQATNGTN---SDSLKSTQDEIQORLEEDRVSNQTFNGVKVLSQDNQMK 144
QY 99 VE 100
: :
Db 145 IQ 146
Search completed: July 16, 2003, 12:42:59
Job time : 24 secs

THIS PAGE BLANK (USPTO)